



Université
de Toulouse

THÈSE

En vue de l'obtention du DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Pathologie, Toxicologie, Génétique et Nutrition

Présentée et soutenue par :

Anne Costard

le : 10 novembre 2010

Titre :

Optimisation de modèles permettant de réduire l'incidence d'une maladie par
l'introduction de nouveaux critères de sélection liés à de(s) gène(s) de
résistance chez les animaux domestiques

JURY

Laurence Moreau (Examinatrice)

Didier Concordet (Président du jury)

Thomas Schiex (Examineur)

Florence Phocas (Rapporteur)

Miguel Toro (Rapporteur)

Ecole doctorale :

Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Unité de recherche :

Station d'Amélioration Génétique des Animaux -INRA

Directeur(s) de Thèse :

Jean-Michel Elsen

Rapporteurs :

Florence Phocas

Miguel Toro

Optimisation de modèles permettant de réduire l'incidence
d'une maladie par l'introduction de nouveaux critères de
sélection liés à de(s) gène(s) de résistance chez les animaux
domestiques

Anne Costard

14 décembre 2010

Remerciements

Une thèse est une aventure personnelle qui représente quatre années de la vie, quatre années difficiles, jalonnées de déceptions, de satisfactions, de rencontres scientifiques, de discussions, de débats, parfois de mauvaise humeur. Une thèse est donc aussi une aventure collective, humaine. Si cette thèse a pu être effectuée dans de bonnes conditions, et déboucher sur des résultats passionnants, c'est grâce au concours de personnes que je souhaite remercier ici.

Je remercie Christèle Robert directrice de la Station d'Amélioration Génétique des Animaux de l'Inra d'Auzeville ainsi que les anciens directeurs Eduardo Manfredi et Alain Ducos pour les moyens qu'ils ont mis à ma disposition afin de réaliser ce travail.

Je tiens à remercier les membres du jury ainsi que les rapporteurs pour l'intérêt qu'ils ont eu pour mon travail.

Je remercie également Jean Michel Elsen pour son encadrement tout au long de ma thèse, pour ses conseils et ses compétences scientifiques qui m'ont beaucoup aidés.

Mes remerciements vont également à tous les gens de la station qui m'ont accueilli chaleureusement.

Je tiens à remercier également Carine Naudinat pour toute l'aide qu'elle m'a apportée pendant ces quatre années.

Un grand merci à Mylène, Ingrid et Kin-Anh pour m'avoir soutenue moralement dans les bonnes et mauvaises périodes pendant ces quatre années, doctorantes qui ont soutenu leur thèse pour qui le pire est derrière et le meilleur est à venir,

Je tiens à remercier tout particulièrement Bérengère pour son soutien, pour la décoration de mon bureau et de m'avoir motivée pour la fin de ma thèse (partie la plus difficile...).

Je tiens également à remercier l'équipe des informaticiens et des électroniciens pour leurs apéros et leurs bonnes humeurs ainsi que les caliméros (Hugues, Simon et Julien) pour les petites pauses café et cigarettes où la rigolade était toujours de la partie.

J'adresse toute mon affection et ma reconnaissance à ma maman ainsi qu'à mon mari qui m'ont soutenu tout au long de ces quatre années et qui ont toujours cru en moi.

Je tiens à dédier ce travail à MARIA, ma grand mère paternelle qui m'avez inculqué des valeurs fondamentales de hargne, de volonté et d'humanité sans lesquelles je n'aurais sans doute pas pu réussir ces études. Tu es malheureusement partie trop tôt pour voir cet aboutissement, mais j'espère que là où tu es, tu es fier de moi.

Je dédie également cette thèse à ma MAMAN pour m'avoir donné les capacités et les armes nécessaires pour affronter cette expérience et celles qui vont suivre. Tu m'as toujours soutenu, aidé, et tu m'as donné les moyens de faire des études et de les mener à bien, je t'en serai toujours reconnaissante.

Je dédie également cette thèse à FREDO, mon mari. Toi qui partage ma vie, tu m'as supporté pendant ces 4 années et je le sais je n'ai pas toujours été facile! Mais ton soutien et ton amour m'ont permis de faire face à toutes les difficultés. Alors cette thèse, c'est aussi la tienne

Enfin je dédie cette thèse à DORIAN, mon fils. Ta merveilleuse arrivée lors de ce travail m'a redonné la motivation nécessaire pour le finir et surtout m'a permis de prouver qu'il était possible d'effectuer une thèse tout en ayant eu un enfant. Tu es un rayon de soleil, celui qui embellit mon coeur chaque jour que Dieu fait. Avec toi, le monde est une merveille, un monde où il n'existe que le bonheur.

Résumé

Les moyens mis en oeuvre chez les animaux domestiques pour lutter contre les maladies dépendent de leur vitesse de propagation, de leur virulence et surtout de leur capacité à passer la barrière d'espèce. Certaines maladies entraînent des pertes économiques importantes chez l'animal (des centaines de milliers de vaches ont été abattues pendant l'épidémie de BSE). Bien plus, dans le cas des zoonoses, maladies transmissibles à l'homme, il convient d'appliquer des mesures de lutte extrêmement onéreuses s'organisant dans les élevages mais aussi à l'échelle des populations. Plusieurs stratégies de prévention et/ou de traitement des maladies infectieuses sont possibles : les traitements chimiques, la vaccination, la désinfection des bâtiments d'élevage, la conduite des troupeaux, le contrôle sanitaire et la traçabilité des produits alimentaires, l'abattage partiel ou total des troupeaux infectés. Ces mesures peuvent être très onéreuses et parfois peu efficaces.

Ces maladies peuvent être mono ou multifactorielles. Quand une maladie est en partie régulée par une prédisposition génétique, il est possible de choisir des reproducteurs les plus résistants et diminuer ainsi l'incidence de la maladie. Si cette prédisposition est monogénique, ou pour le moins sous l'influence d'un gène à effet majeur, comme par exemple dans la tremblante ovine, la sélection assistée par gène peut être une stratégie complémentaire aux stratégies déjà existantes. Mais cette sélection ne peut se faire indépendamment de celle des caractères traditionnels de production qui ont permis d'améliorer par exemple la quantité de lait en races laitières ou le nombre et la croissance des animaux en races bouchères. La question qui se pose alors est d'arriver le plus rapidement possible à une incidence quasi nulle de la maladie en créant une population essentiellement porteuse des allèles de résistance au gène sélectionné tout en évitant de perdre le gain génétique acquis sur les caractères traditionnels de production.

L'objectif de cette thèse a été de répondre à cette question en proposant une approche paramétrique permettant de modéliser et d'optimiser la stratégie de sélection. Le paramétrage du modèle permet de décrire de manière détaillée les caractéristiques du/des gène(s) de résistance sélectionné(s) et celles du schéma de sélection sur les caractères de production, souvent complexes chez les animaux domestiques. Les sorties du modèle sont l'évolution des distributions des caractères de production (moyennes, variabilités) et des fréquences des allèles de résistance / sensibilité. Afin de prendre en compte le processus de sélection, une modélisation dynamique a été utilisée. Des approches déterministe et stochastique ont été développées et ont été optimisées à l'aide d'un algorithme génétique.

La stratégie optimale difficile à promouvoir sur le terrain a été alors considérée comme référence pour la mise en place de stratégies de sélection acceptables et efficaces. Il est possible ainsi d'évaluer différents scénarios concernant par exemple l'organisation du génotypage (pour le gène de résistance) dans la population sélectionnée, le choix des conjoints selon leur génotype, ainsi que la méthode de diffusion des reproducteurs vers les élevages de production. La méthodologie développée lors de ce travail a été appliquée au cas de la tremblante du mouton, un exemple possible parmi d'autres maladies. Différentes stratégies proches de l'optimale ont été combinées à un modèle épidémiologique afin d'analyser l'impact de ces dernières sur l'évolution de la maladie.

MOTS CLEFS : gène(s) de résistance aux maladies, modélisation, optimisation, sélection, caractères d'intérêt zootechnique, incidence, animaux domestiques.

Summary

The resources to be used in domestic animals to fight against the diseases depend on their diffusion speed, their virulence and especially their ability to pass the species barrier. Some diseases cause significant economic losses in animals (hundreds of thousands of cows were culled during the BSE epidemic). Moreover, in the zoonotic case, diseases transmissible to humans, measures to be implemented are generally extremely expensive both at the farm and population levels. Several strategies for the prevention and / or treatment of infectious diseases are possible : chemical treatments, vaccination, disinfection of livestock buildings and herd management, health monitoring and traceability of food products, or partial or full slaughtering of herds . These measures can be very costly and sometimes ineffective. These diseases can be mono or multifactorial. In the case of genetic diseases (most often single gene such as sheep scrapie), gene-assisted selection can be a complementary strategy to existing strategies. But this selection can not be done independently of the selection of the traditional production traits such as the quantity of milk in dairy breeds or the number and the growth of animals in beef breeds. The question that arises is how to reach as quickly as possible an almost zero incidence of the disease by creating a population displaying a high proportion of the resistant allele at the gene of interest while minimizing the loss of genetic progress on traditional production traits. The objective of this thesis was to answer this question by proposing a parametric approach to model and optimize the selection strategy. The parametrization of the model allowed to describe in detail the characteristics of the gene (s) selected for resistance (s) and those of the breeding scheme on production traits, often complex in domestic animals. The model outputs were the evolution of distributions of production traits (means, variability) and allele frequencies of resistance / sensitivity. To take into account the selection process, a dynamic model was used. Deterministic and stochastic approaches have been developed and have been optimized using a genetic algorithm. The optimal strategy may be difficult to apply. The optimal strategy may be taken as a reference for the development of selection strategies. It is thus possible to evaluate different scenarios such as the genotyping organization (for the resistance gene) in the selected population, the choice of the mating according to their genotype, the method of dissemination of reproducers to the production farms. The methodology developed in this work has been applied to the scrapie case, a possible example among other diseases. Different strategies close to optimal have been combined with an epidemiological model to analyze their impact of the spread of the disease.

KEYWORDS : selected gene for resistance , modelization, optimization, selection, traditional production traits, incidence, domestic animals.

Table des matières

Introduction	6
I Bibliographie	8
1 Résistance aux maladies	9
1.1 Description de la résistance aux maladies	9
1.2 Maladies à prion	13
1.2.1 L'encéphalopathie Spongiforme Bovine (ESB)	13
1.2.2 La maladie de Creutzfeld-Jakob (MCJ)	13
1.2.3 La Tremblante	14
1.2.4 Transmission inter-espèce	18
1.2.5 Nouvelle forme : la tremblante atypique	18
1.3 Description des méthodes de sélection pour la résistance aux maladies	18
1.4 Le programme français d'éradication de la tremblante	20
1.5 Conclusion	21
2 Modélisation de la sélection	23
2.1 Introduction	23
2.2 Objectif et critère de sélection	23
2.3 Indexation	24
2.3.1 Principes généraux de l'indexation	24
2.3.2 Différentes méthodes d'indexation	24
2.4 Réponse à la sélection et progrès génétique	25
2.4.1 Intervalle de génération	25
2.4.2 Variabilité génétique	26
2.4.3 Intensité de sélection	26
2.4.4 Précision de la sélection	26
2.5 Schéma de sélection	26
2.6 Sélection assistée par marqueurs	29
2.7 Optimisation de la sélection	31
3 Optimisation	32
3.1 Modèle	32
3.2 Méthodes d'optimisation	33
3.2.1 Méthode de Lagrange	33
3.2.2 <i>Sequential Quadratic Programming SQP</i>	34

3.2.3	Programmation semi-définie SDP	35
3.2.4	Programmation dynamique	35
3.2.5	Recuit simulé	36
3.3	Les algorithmes génétiques	38
3.4	Les différents opérateurs des algorithmes génétiques	41
3.4.1	Opérateur de sélection	41
3.4.2	Opérateur de reproduction	42
3.4.3	Les paramètres de dimensionnement	44
3.5	Les améliorations classiques	45
3.5.1	Scaling	46
3.5.2	Sharing	46
3.5.3	Parallélisation	47
3.6	Convergence	50
3.7	Conclusion	51
II	Modèle déterministe	52
4	Modèle déterministe	53
4.1	Résumé	53
4.2	Article	55
III	Modèle stochastique	84
5	Modèle stochastique	85
5.1	Résumé	85
5.2	Article	87
IV	Modèle épidémio-génétique	117
6	Combinaison Modèle génétique-Modèle épidémiologique	118
6.1	Résumé	118
6.2	Article	119
V	Conclusion	135
	Conclusion	136

Liste des tableaux

1.1	Différentes maladies régulées par l'effet d'un gène majeur	11
1.2	Quelques exemples de maladies génétiques humaines	12
1.3	Récapitulatif des risques d'infection de l'animal selon son génotype	16
1.4	Fréquences des allèles du gène PrP dans les races ovines françaises (Palhière, 2002)	17
1.5	Fréquences des allèles du gène PrP dans les troupeaux ovins français (agneaux et agnelles) en 2008	22
3.1	Algorithme du recuit simulé	37
3.2	Exemple de représentation d'un individu	38
3.3	Exemple de sélection par rang (source : http://sis.univ-tln.fr)	42

Table des figures

1.1	Mouton atteint de la tremblante	14
1.2	Caractéristiques au niveau tissulaire et neuronal de la tremblante	15
2.1	Intervalle de génération entre 2 générations succesives t et t+1	26
2.2	Une génération de sélection (Minvielle (1998))	27
2.3	Principe de la sélection assistée par marqueurs	29
3.1	Fonctionnement général des algorithmes génétiques (source : www.recherche.enac.fr)	40
3.2	Exemple de sélection par roulette de wheel (source : http://sis.univ-tln.fr) . .	41
3.3	Exemple du principe du croisement à un point	43
3.4	Exemple du principe du croisement uniforme	44
3.5	Exemple de risque de reproduction d'un seul individu	45
3.6	Principe du Sharing	46
3.7	Fonctionnement du modèle en îlots	47
3.8	Fonctionnement du modèle à population distribuée	48
3.9	Schéma du Maître-Esclave synchrone	49
3.10	Schéma du Maître-Esclave asynchrone	50

Introduction

Introduction

Les moyens mis en oeuvre chez les animaux domestiques pour lutter contre les maladies dépendent de leur vitesse de propagation, de leur virulence et surtout de leur capacité à passer la barrière d'espèce. Certaines zoonoses entraînent pour les reproducteurs, des pertes économiques non négligeables. Bien plus, le fait que l'Homme soit menacé conduit à l'application de mesures de lutte extrêmement onéreuses. Celles-ci s'organisent dans chaque élevage mais aussi à l'échelle des populations. Plusieurs stratégies de prévention et/ou de traitement des maladies infectieuses sont possibles, souvent très onéreuses et parfois peu efficaces. Dans ce contexte, la stratégie génétique qui exploite l'existence d'une variabilité génétique de la résistance/sensibilité à une maladie particulière, peut intervenir en complémentarité des stratégies déjà existantes.

La sélection pour la résistance à une maladie ne peut cependant pas se faire indépendamment de celle des caractères traditionnels de production. Il s'agit donc d'arriver le plus rapidement possible à une incidence quasi nulle de la maladie en créant une population essentiellement porteuse des allèles de résistance au(x) gène(s) sélectionné(s) en minimisant l'impact de cette sélection sur le gain génétique acquis pour les caractères traditionnels de production.

L'objectif de ce travail de thèse est de proposer des approches paramétriques permettant de modéliser et d'optimiser cette stratégie de sélection. Le paramétrage des modèles décrit de manière détaillée les caractéristiques du gène de résistance sélectionné et celles du schéma de sélection sur les caractères de production.

Une revue bibliographique concernant la résistance aux maladies, la modélisation de la sélection ainsi que les méthodes d'optimisation font l'objet du premier chapitre. Afin de

prendre en compte le processus de sélection, une modélisation dynamique a été utilisée. Des approches déterministe et stochastique ont été envisagées (partie 2 et 3). Ces modélisations permettent d’obtenir des stratégies optimales et d’évaluer différents scénarios (organisation du génotypage pour le gène de résistance dans la population sélectionnée, choix des conjoints selon leur génotype, méthode de diffusion des reproducteurs vers les élevages de production).

Chez les animaux domestiques, la recherche de gènes de résistance aux maladies est devenue un sujet de recherche prioritaire en génétique. Le cas de la tremblante du mouton est un exemple type. Il a été démontré qu’un gène (*PrP*) exerce un contrôle très important sur la sensibilité à cette maladie. Différents modèles épidémiologiques concernant la tremblante ont été développés afin de simuler différents scénarios de transmission et de comprendre l’évolution de cette maladie. Dans ce travail, nous avons cherché à quantifier si l’optimisation de la sélection simultanée sur le gène PrP et sur un caractère de production avait un réel impact sur l’évolution de la maladie. Le modèle déterministe développé dans le chapitre 2 a été couplé avec un modèle dynamique épidémiologique afin d’étudier l’impact des différentes stratégies de sélection sur l’évolution de la maladie (Partie 4).

Première partie

Bibliographie

Chapitre 1

Résistance aux maladies

1.1 Description de la résistance aux maladies

La résistance génétique aux maladies chez les animaux domestiques a suscité un très grand intérêt depuis de nombreuses années. En effet, les maladies des animaux domestiques entraînent une perte importante au niveau de la production et un risque pour l'Homme. La résistance aux maladies s'exprime de différentes manières. La résistance/la sensibilité peut être totale (mortalité ou survie) ou partielle (intensité plus ou moins forte des signes cliniques). D'autre part, certains animaux « porteurs de la maladie » peuvent maintenir leur productivité tout en étant infecté, ce phénomène est appelé la résilience. Les porteurs sains, qui ne sont pas identifiés comme étant « malades », représentent un danger potentiel pour les autres animaux de la même espèce et pour les autres espèces. Ils constituent un réservoir de l'agent pathogène et sont un obstacle à son élimination. Les maladies peuvent être mono ou multifactorielles. Parmi les facteurs figure souvent une plus ou moins grande prédisposition génétique. On distinguera les maladies génétiques, le plus souvent monogénique, et les maladies nécessitant une infection par des bactéries, des virus ou des prions ou une infestation par des parasites. Les premières sont associées au dysfonctionnement d'un gène particulier. Si la sensibilité à la maladie est liée à un allèle dominant alors cette sensibilité se retrouve chez tous les descendants d'un individu porteur de cet allèle. Si la sensibilité est associée à un allèle récessif, alors les descendants d'un individu porteur de cet allèle récessif seront porteurs de cette sensibilité mais ne développeront pas la maladie. La sensibilité génétique à une maladie infectieuse ou parasitaire peut être le fait d'un (monogénique) ou de plusieurs (oligo ou multigéniques) gènes.

Pour lutter contre les maladies, différents moyens sont utilisés : traitements chimiques (antibiotiques, bactéricides, vaccins), désinfections des bâtiments d'élevage, abattage partiel ou total des troupeaux infectés. Ces diverses mesures s'avèrent parfois non satisfaisantes et/ou très coûteuses et/ou difficiles de mise en place. Les traitements peuvent avoir des impacts négatifs sur l'environnement et sur la sécurité alimentaire. La résistance de certains agents pathogènes face aux traitements évolue. Pour ces raisons, l'étude de la variabilité génétique pour améliorer la résistance aux maladies offre un outil complémentaire pour contrôler ces maladies. Différentes méthodes existent pour étudier la variabilité de la résistance aux maladies selon qu'elle soit polygénique, oligogénique ou monogénique.

Dans le cas polygénique, il convient d'estimer les héritabilités d'indicateurs de résistance et leurs corrélations. Ces estimations se font à l'aide de modèles statistiques prenant en compte les valeurs des caractères mesurés, l'apparementement entre individus et les effets des milieux.

Dans le cas oligogénique, qui suppose que quelques gènes ont un effet important sur le caractère de résistance, la recherche de ces gènes à l'aide de marqueurs moléculaires est réalisable grâce aux cartes génétiques. Ces informations permettent de détecter des zones chromosomiques porteuses de gènes liés aux caractères d'intérêts : les Quantitative Trait Loci (QTL).

Dans le cas monogénique, la variabilité du caractère de sensibilité est due à l'effet d'un gène « majeur ». Le principe de localisation est identique à celui précédemment cité ou peut consister en une analyse d'association entre la sensibilité et le génotype en un gène candidat.

Dans les tableaux 1.1 et 1.2, quelques exemples de maladies génétiques dans différentes espèces animales et chez l'homme sont répertoriés. Cette liste non exhaustive ainsi que différentes sites comme orpha.net ou celui de « NCBI » montrent que le nombre de maladies à composante génétique sont très nombreuses.

Dans ce travail, nous nous sommes particulièrement intéressés aux maladies à prion et plus particulièrement à la tremblante ovine. En effet, cette maladie présente un intérêt majeur du point de vue vétérinaire ainsi qu'au niveau de la compréhension de certaines maladies humaines dont elle se rapproche.

TABLE 1.1 – Différentes maladies régulées par l'effet d'un gène majeur

Maladie	Espèces	Caractéristiques de la maladie	Déterminisme génétique	Références
Ichtyose foetale	Bovine (Chianina)	Maladie de la peau qui se caractérise par une kératinisation exagérée sur l'ensemble du corps aboutissant à la formation d'une sorte de carapace totale particulièrement épaisse qui durcit et qui finit par se craqueler se fendiller.	La mutation responsable de cette maladie affecte le gène ABCA12	Charlier et al. (2008)
dystonie musculaire congénitale CMD1, CMD2	Bovine	Les veaux atteints par la première forme (CMD1) ont des raidissements brusques et ensuite sont dans l'incapacité de plier leur jambe pour se coucher. Pour la seconde forme, les animaux ont des contractures musculaires généralisées.	La mutation responsable de cette maladie affecte le gène ATP2A1(CMD1) et SLC6A5 (CMD2)	Charlier et al. (2008)
Complex vertébral malformation CVM	bovine	Maladie qui se traduit par des malformations de la colonne vertébrale, une déviation des articulations des membres et biologiquement par des avortements et naissance avant terme de veaux non viables.	La mutation responsable de cette maladie affecte le gène SLC35A3	Agerholm et al. (2001), Charlier et al. (2008)
CLAD (Canine Leukocyte Adhesion Deficiency)	Canine (chiot Setter Irlandais)	Maladie génétique due à une déficience des leucocytes entraînant des infections variées puis la mort dans la première année de leur vie	La mutation responsable de cette maladie affecte le gène ITGB2.	Kijas et al. (2000), Kijas et al. (1999)
BLAD (Bovine Leukocyte Adhesion Deficiency)	Bovine (race Holstein)	Maladie génétique où les animaux atteints sont sensibles à de simples infections et meurent dans l'année de leur naissance de pneumonie, septicémie, méningite.	La mutation responsable de cette maladie affecte le gène CD18.	Shuster et al. (1992), Gerardi (1996)
EBJ(Epidermolyse Bulleuse Junctionnelle)	Equine (chevaux de trait), Canine	Maladie génétique qui se traduit par un défaut d'adhérence de la couche basale de l'épiderme et du derme. Cette maladie se caractérise par de multiples lésions entraînant des surinfections puis une septicémie fatale en quelques jours.	La mutation responsable de cette maladie affecte le gène LAMC2.	Milenkovic et al. (2003), Guaguère et al. (2003)
Dermatophilose	Bovine	Maladie infectieuse, virulente, inoculable et contagieuse. Elle se caractérise par une épidermite exsudative et croûteuse. Cette maladie diminue la productivité et accroît le taux de mortalité.	La mutation responsable de la sensibilité à cette maladie affecte le gène BoLA-DRB3-DQB	Maillard et al. (2003)
Maladie de Marek	Poulet	La perte de poids peut être le seul symptôme évident avant la mort. Une paralysie complète ou partielle de différentes parties du corps entraîne : ailes pendantes, boiterie, difficulté respiratoire ou attitudes anormales.	B21	Hansen et al. (1967)
Leucose	Bovine	Une maladie virale qui se manifeste par des tumeurs.	CMH	Lewis (1986)
Escherichia Diarrhea	Porcine	Cette maladie apparaît chez des porcelets dans la première semaine. Les porcelets meurent très vite avec une diarrhée extrêmement fluide.	K88	Gibbons et al. (1977)

TABLE 1.2 – Quelques exemples de maladies génétiques humaines

Maladie	Caractéristiques	Gène en cause	Références
Cadasil	migraines,accidents vasculaires cérébraux,troubles psychiques,troubles de la mémoire, des difficultés de déplacement et parfois une perte d'autonomie	NOTCH3	Joutel et al. (1996)
Fièvre méditerranéenne familiale	douleurs abdominales, dans le thorax et dans les articulations. Principale complication de la maladie : insuffisance rénale nécessitant d'être dialysé.	MEFV	Notarnicola et al. (2002)
maladie de Bietti	rétrécissement de champ visuel de manière progressive	CYP4V2	Li et al. (2004)
Maladie de Caffey	Maladie infantile osseuse caractérisée par une augmentation du volume de certaines parties du squelette accompagnée de douleurs musculaires et d'inflammations	COL1A1	Gensure RC (2005)
Maladie de Paget	Hypertrophie osseuse	SQSTM1, TNFRSF11B	Collet et al. (2007)
Maladie de Parkinson, forme familiale	Dégénérescence de certains neurones provoquant des troubles des mouvements	MAPT	Ghetti et al. (2003)
Maladie de Rendu-Osler	Malformations artério-veineuses cutanées ou viscérales entraînant des saignements du nez, des malformations vasculaires pulmonaires, cérébrales, hépatiques, vertébrales et intestinales	ENG (HHT1) / ACVRL1 (HHT2) / SMAD4 (Juvenile HHT)	Shovlin et al. (2000)
Mucoviscidose	Problèmes respiratoires (toux, expectoration), digestifs (ballonnements, douleurs abdominales parfois diarrhée) , mauvaise croissance.	CFTR	Fanen & Hasnain (2001)
Nanisme diastrophique	Raccourcissement des membres avec un crâne de taille normale, déformation de la colonne vertébrale et contractures articulaires	SLC26A2	Rossi (2001)
Sclérose mésentérique diffuse	Insuffisances rénales justifiant un programme de dialyse-transplantation	WT1	Niaudet & Gubler (2007)
Syndrome de BIDS	cheveux épars et cassants, retard mental, hypofertilité et nanisme	C7ORF11	Nakabayashi et al. (2005)
Syndrome de Cohen	déficience intellectuelle, troubles de la vue (strabisme), obésité	COH1	Kolehmainen et al. (2003)
Syndrome de Larsen	Déboitements des articulations(genoux, poignet,épaule)	FLNB	Zhang et al. (2006)
syndrome de l'ostéoporose pseudogliome	Fragilité osseuse	LRP5	Gong et al. (2001)

1.2 Maladies à prion

Les maladies à prion sont des pathologies de l'encéphale associées à la présence de particules infectieuses protéiniques couramment appelées prions. Le nom scientifique de ces maladies à prion est Encéphalopathies Spongiformes Subaiguës Transmissibles (ESST). Ces maladies à prion touchent de nombreuses espèces : bovins (encéphalopathie spongiforme bovine), ovins (tremblante), humains (maladie de Creutzfeldt-Jakob).

L'agent infectieux responsable de ces maladies est très particulier. En effet, il résiste aux procédés « habituels » de destruction : ultraviolets, radiations ionisantes, chaleur humide ou sèche (130°C), enzymes (8,9). En 1982, Stanley Prusiner (Prusiner (1982)) a montré que l'agent infectieux est une protéine sans acides nucléiques qu'il a appelé « prion ». Les prions proviennent d'un changement de conformation 3D de la protéine PrP ou PrPc (cellulaire) en protéine scrapie (PrPsc) ou résistante (PrPres). La protéine PrPc est relativement présente dans le système nerveux central, le tissu lymphoïde et le tube digestif. La PrPres a pour caractéristiques de transmettre, par des interactions protéine-protéine, ses propriétés à la protéine « normale ». L'accumulation de la PrPres est corrélée avec la destruction des cellules nerveuses et l'infectiosité des tissus porteurs.

1.2.1 L'encéphalopathie Spongiforme Bovine (ESB)

Cette maladie plus connue sous le nom de la maladie de la « vache folle » est apparue en 1986 (Wells et al. (1987)) au Royaume Uni. Environ 220 000 cas ont été détectés dont 90% en Grande Bretagne. Les autres cas sont apparus en Suisse, Belgique, Portugal, France et Allemagne et plus tardivement au Japon et au Canada. Cette maladie a été principalement présente au Royaume Uni. Les premiers symptômes externes apparaissent à l'âge adulte (4-5 ans en moyenne). Cette maladie fatale affecte le cerveau et la moelle épinière et se traduit par des troubles du comportement, une baisse de la production laitière, perte de poids, troubles locomoteurs. En 1993, Foster *et al* (Foster et al. (1993)) ont prouvé la transmissibilité de l'ESB aux ovins et aux caprins. De plus, des études ont montré que les cas de la nouvelle variante de la maladie de Creutzfeldt-Jakob, apparue pour la première fois en 1996, auraient pour origine le même agent que celui responsable de l'ESB.

1.2.2 La maladie de Creutzfeld-Jakob (MCJ)

La maladie de Creutzfeld-Jakob (MCJ), qui affecte l'homme, a été décrite pour la première fois en 1920. Cette maladie rare et fatale affecte le système nerveux central. Trois formes de cette maladie sont connues :

- sporadique (85% des cas),
- familiale héréditaire (10 à 15% des cas),
- iatrogénique (<5% des cas) c'est-à-dire que l'individu est contaminé par un processus opératoire : utilisation d'hormones ou greffes de tissus cérébraux issus de cadavres malades ou encore par utilisation d'instruments chirurgicaux contaminés.

Cette maladie touche en général des personnes âgées entre 50 et 75 ans avec une évolution très rapide (4,5 mois). Elle se caractérise par des troubles de l'équilibre, de la sensibilité puis par une démence. La nouvelle variante de la maladie de Creutzfeld-Jakob (nvMCJ) atteint des patients plus jeunes (27 ans en moyenne) et évolue plus lentement (14 mois)

1.2.3 La Tremblante

Cette maladie est aussi connue sous son terme anglophone « scrapie ». Elle a été diagnostiquée en Europe chez les ovins et les caprins en 1732 (Comber (1772)). La répartition de cette maladie est mondiale à l'exception notable de l'Australie et de la nouvelle Zélande.

Caractéristiques de la tremblante

Dans la majorité des cas, les symptômes de la tremblante sont :

- troubles de la motricité (tremblements, perte d'équilibre, incoordination des mouvements) ;
- hyper sensibilité (réaction exagérée au contact et au bruit) ;
- troubles du comportement (crainte, agressivité, torpeur) ;
- démangeaisons ;
- perte de poids.

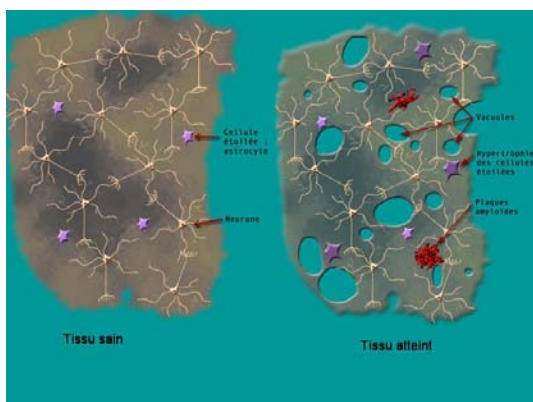
Cette maladie se caractérise par une période d'incubation assez longue de quelques mois à plusieurs années (Sigurdsson (1954)). L'évolution clinique est assez variable (1 à 12 mois) mais inévitablement mortelle. Cette évolution peut se résumer en quatre étapes. Premièrement, l'animal s'isole. Un prurit apparaît de la tête à la région dorsolombaire ce qui provoque des zones dépilées (Figure 1.1). L'animal est très excité et tremble. Deuxièmement, les signes



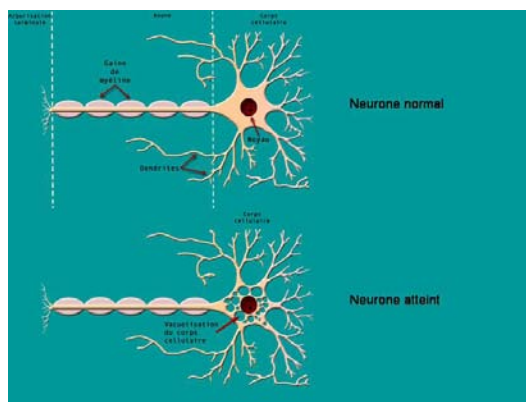
FIGURE 1.1 – Mouton atteint de la tremblante

décrits précédemment s'aggravent. On voit également apparaître le développement d'une incoordination locomotrice (démarche hésitante, trébuchements, chutes). L'état général de l'animal se dégrade et des signes d'anxiété se développent. Troisièmement, les troubles moteurs sont de plus en plus importants et des lésions nerveuses apparaissent. L'animal reste souvent couché. Quatrièmement, l'animal ne peut plus se tenir debout puis meurt. La tremblante peut être diagnostiquée par l'observation des symptômes cités ci-dessus. Le diagnostic est confirmé par un examen histopathologique permettant de mettre en évidence des lésions spongiformes du cerveau caractéristiques de ce groupe de maladies. A l'aide de l'étude microscopique de coupes de cerveau de mouton, on peut observer au niveau tissulaire des vacuoles, espaces vides laissés par la disparition de neurones (Figure 1.2a). Ces vacuoles montrent la dégénérescence neuronale. Au niveau neuronal, on observe la vacuolisation du corps cellulaire (Figure 1.2b). Ce diagnostic plus poussé est nécessaire afin de connaître l'origine de l'agent infectieux. En effet, les signes cliniques ne permettent pas la différenciation de l'origine de l'agent infectieux

entre le tremblante et l'ESB et ne peut se faire que par un examen histopathologique. Il est nécessaire de savoir l'origine de la maladie puisque contrairement à la tremblante, l'ESB comporte un risque pour l'homme.



(a) Schéma de tissus sains et de tissus atteints



(b) Schéma d'un neurone sain et d'un neurone atteint

FIGURE 1.2 – Caractéristiques au niveau tissulaire et neuronal de la tremblante

Mode de transmission de la tremblante

Dès 1936, Cuillé et Chelle (Cuille & Chelle (1936)) ont montré la transmissibilité de cette maladie. Cette transmissibilité se fait principalement par voie orale. Ils sont parvenus également à la transmettre à une chèvre montrant ainsi que le franchissement de la barrière d'espèce était possible. Dans les années 60, David Haig est parvenu à transmettre la tremblante ovine à la souris par inoculation intracérébrale. La contamination du mouton par la consommation de placentas de brebis infectées a été démontrée (Sarradin et al. (1996)), ainsi que celle par le pâturage en zone herbeuse contaminée par l'agent de la tremblante (Sarradin et al. (1996)). La transmission de l'infection à la souris a été réussie à partir de broyats d'acariens du foin (Wisniewski et al. (1996)), ce qui laisse supposer que les fourrages pourraient servir de réservoirs.

Grâces aux résultats obtenus par la sélection de reproducteurs résistants à une inoculation expérimentale (Davies & Kimberlin (1985)), l'existence d'un contrôle génétique de la sensibilité à la tremblante a été démontrée. L'effet d'un gène majeur, le gène Sip (« *Scrapie incubation period* »), sur le contrôle de la durée d'incubation de la tremblante ovine a été prouvé en 1968 (Dickinson et al. (1968)). Une forte liaison génétique a été montrée entre les loci de Sip et du gène codant pour la protéine prion PrP. Le gène PrP est polymorphe en divers codons. Le polymorphisme aux codons 136, 154 et 171 est lié à des différences de sensibilité à la tremblante expérimentale et naturelle (Goldmann et al. (1991), Clouscard et al. (1995), Elsen et al. (1999), Hunter et al. (1996)); 25 allèles de ce gène sont répertoriés parmi lesquels quatre allèles sont fortement associés à la variabilité génétique de la sensibilité à la tremblante : ARR, ARQ, AHQ, VRQ (Belt et al. (1995)).

Différentes combinaisons des quatre principaux allèles ont été recensées et ont été classées

TABLE 1.3 – Récapitulatif des risques d'infection de l'animal selon son génotype

Génotype	Risques
ARR/ARR	Très faible risque de contamination de l'animal et faible risque sur ses descendants de première génération
ARR/AHQ AHQ/AHQ	Faible risque de contamination de l'animal et sur ses descendants de première génération
ARR/ARQ AHQ/ARQ	Risque sur ses descendants de première génération
ARQ/ARQ ARR/VRQ AHQ/VRQ	Risque de contamination de l'animal et risque important pour ses descendants de première génération
ARQ/VRQ VRQ/VRQ	Risque important de contamination pour l'animal

selon le niveau de risque de contamination de l'animal et de ses descendants (Tableau 1.3). L'allèle VRQ confère une hypersensibilité à la tremblante ovine. Les animaux homozygotes VRQ/VRQ contaminés par le prion développent rapidement la maladie. La diffusion du prion se fait très rapidement après la contamination par les voies nerveuses et lymphatiques. Les ovins porteurs de l'allèle ARQ sont relativement sensibles à la maladie. Quant à l'allèle AHQ, il détermine une sensibilité intermédiaire. Enfin l'allèle ARR confère une grande résistance à la maladie. Les animaux hétérozygotes ARR/XXX sont très peu sensibles à la tremblante. Les fréquences des ces quatre allèles dans la population ovine française, estimées avant la mise en place d'un plan de sélection pour la résistance à la tremblante, sont récapitulées dans le tableau 1.4.

TABLE 1.4 – Fréquences des allèles du gène PrP dans les races ovines françaises (Palhière, 2002)

Races	ARR	AHQ	ARQ	VRQ
Ile de France	0.687	0.000	0.146	0.167
Rava	0.430	0.007	0.528	0.035
Charmoise	0.313	0.030	0.495	0.162
Berrichon du Cher	0.805	0.063	0.105	0.026
Est à laine mérinos	0.159	0.066	0.774	0.000
Limousine	0.406	0.000	0.594	0.000
Bleu du Maine	0.700	0.000	0.050	0.250
Texel	0.270	0.050	0.590	0.090
Suffolk	0.704	0.000	0.281	0.015
Hampshire	0.602	0.005	0.379	0.015
Mérinos d'Arles	0.359	0.025	0.591	0.025
Préalpes du Sud	0.441	0.000	0.559	0.000
Bizet	0.632	0.019	0.349	0.000
Noir du Velay	0.227	0.000	0.736	0.036
Lacaune Gebro	0.566	0.010	0.278	0.146
Lacaune Ovitest	0.400	0.030	0.505	0.065
Blanc du massif Central	0.250	0.050	0.621	0.079
Causses du Lot	0.151	0.174	0.604	0.071
Grivette	0.441	0.000	0.544	0.015
Mouton Vendéen	0.163	0.000	0.822	0.015
Rouge de l'ouest	0.667	0.000	0.250	0.083
Tarasconnaise	0.325	0.010	0.660	0.005
INRA 401	0.360	0.071	0.458	0.111
Lacaune Lait	0.545	0.016	0.427	0.012
Basco Béarnaise	0.399	0.000	0.601	0.000
Corse	0.470	0.039	0.487	0.003
Manech Tête Noire	0.496	0.012	0.488	0.004
Manech Tête Rousse	0.167	0.008	0.806	0.019

1.2.4 Transmission inter-espèce

La tremblante a été transmise expérimentalement à différentes espèces (souris, hamster, bovins, caprins) par voie intra-cérébrale ou orale. Les seules espèces contaminées, dans des conditions naturelles par la tremblante sont la chèvre et le vison d'élevage.

Contrairement à l'ESB, l'agent de la tremblante ne semble pas passer facilement les barrières d'espèce, en particulier celle concernant l'espèce humaine. En effet, aucun lien n'a été démontré entre la tremblante et la MCJ. Néanmoins les ovins peuvent être contaminés par l'agent infectieux de l'ESB et développer une forme de « tremblante ». Par conséquent, on ne peut pas écarter totalement le risque de contamination pour l'être humain.

1.2.5 Nouvelle forme : la tremblante atypique

En 1998, une nouvelle souche particulière de tremblante ovine nommée Nor98 est apparue en Norvège (14). A partir de là, d'autres cas discordants ont été détectés en Europe (Buschmann et al. (2004), Gavier-Wide'n et al. (2004), Konold et al. (2006), Onnasch et al. (2004), Orge et al. (2004)). Cette souche est différente de la souche « classique » sur le plan clinique (Konold et al. (2007), Onnasch et al. (2004)) et physiopathologique (Benestad et al. (2003), Le Dur (2005), Simmons et al. (2007)). Des analyses génétiques ont permis de mettre en évidence une mutation sur le codon 141 (Moum et al. (2005)) chez les animaux dits résistants, jugée sans importance pour la tremblante classique. La tremblante atypique a été transmise par voie intracérébrale à des souris et à un ovin (Simmons et al. (2007)). Cette maladie ne semble pas ou peu contagieuse en conditions naturelles (Luhken et al. (2007), AFSSA (n.d.)).

1.3 Description des méthodes de sélection pour la résistance aux maladies

L'étude de la résistance aux maladies est un sujet majeur qui a pris de l'ampleur grâce à la génétique moléculaire. Les tableaux récapitulant certaines maladies génétiques chez les animaux et chez les humains (liste non exhaustive) montrent que les études concernant la résistance aux maladies sont importantes et ne cessent d'évoluer. Ces recherches dans le milieu animal peuvent aider à mieux comprendre certaines maladies humaines qui s'y apparentent. De plus, pour les éleveurs, la découverte de gènes de résistance permet de sélectionner les animaux résistants, d'amoindrir les pertes économiques occasionnées par une épidémie et de réduire les coûts des traitements chimiques. La mise en place de cette sélection contribue donc à garantir la sécurité du consommateur. Plusieurs stratégies existent pour obtenir une population résistante (sélection phénotypique, sélection génotypique).

La sélection sur le phénotype

Cette sélection est basée sur la réponse phénotypique aux risques des maladies. Cette sélection consiste à choisir systématiquement des reproducteurs résistants le mieux aux maladies, le point clé étant le choix d'un critère de résistance pratique (mesurable en conditions non expérimentales et relativement peu coûteux) et héritable.

Un premier exemple concerne les mammites. Le comptage des cellules somatiques dans le lait et les cas cliniques de la maladie ont été utilisés comme indicateurs de la sensibilité à cette maladie. Il a été montré que la variation de ces facteurs avait une composante génétique (Rupp & Boichard (2003)). Ces données sont utilisées chez les bovins (Shook & Schutz (1994)) et chez les ovins (Barillet et al. (2001)). Un deuxième exemple est la sélection des moutons pour la résistance au parasitisme gastro intestinal, basée sur le comptage des oeufs dans les fèces. Ce critère a été démontré comme un moyen efficace de réduction du traitement médicamenteux et de la contamination des pâturages (Woolaston (1992), Morris et al. (2000), Woolaston & Windon (2001), Bishop & Woolliams (2004)).

Dans certaines espèces, des races sont plus ou moins résistantes à certaines maladies. Il est donc intéressant de croiser les races résistantes avec d'autres races plus sensibles mais avantageuses pour d'autres caractères d'intérêt zootechnique. Par exemple, pour la peste porcine africaine, il a été constaté que les porcs sauvages confèrent une résistance à cette maladie alors que les porcs domestiqués sont atteints. Des accouplements ont été effectués entre ces deux races afin de diminuer la sensibilité à cette maladie.

Dans de nombreuses situations, il n'existe pas de critères pratiques ou peu coûteux permettant d'apprécier la résistance d'un sujet à une maladie donnée. La pression infectieuse peut ne pas être contrôlée, le temps d'incubation peut nécessiter plusieurs années, les pertes économiques ou le coût du confinement des animaux artificiellement infectés peuvent être considérables. En outre, cette sélection ou ces croisements peuvent aboutir à la perte de certains allèles et à la fixation d'autres allèles d'une manière irréversible. En effet, la sélection sur le phénotype de résistance entraîne une sélection involontaire des gènes contrôlant divers caractères, situés sur le même chromosome que les gènes liés à ce caractère. Pour toutes ces raisons, il est préférable, à chaque fois que cela est possible, de réaliser une sélection génotypique.

La sélection sur le génotype

Elle est possible quand un ou un nombre limité de gènes est impliqué dans la variabilité de la résistance. Le polymorphisme a été identifié ou au moins finement localisé à l'aide de marqueurs génétiques en déséquilibre de liaison total avec le gène. Dans cette situation, l'identification des individus porteurs du génotype désiré peut se faire avant qu'ils ne soient en âge de se reproduire. Cette stratégie permet un gain de temps considérable, une amélioration de l'efficacité de la sélection et surtout le contrôle plus facile de certains effets indésirables dus à la sélection (diminution de la variabilité génétique, augmentation de la consanguinité). En effet, la réduction du nombre de reproducteurs et leur utilisation démesurée grâce à l'insémination artificielle restreint les différentes origines des futures générations. Cette sélection génétique pour la résistance aux maladies est particulièrement appropriée aux maladies épidémiques. Elle a été mise en place chez certaines espèces. Pour la maladie de Marek (une maladie virale des poules), l'utilisation des vaccins a accru la virulence de la maladie. La sélection génotypique pour la résistance est apparue nécessaire dans les systèmes de production des volailles. La sélection, basée sur les allèles B21 spécifiques au sein du complexe majeur d'histocompatibilité (CMH) (Bacon (1987)), a été utilisée afin de faciliter la lutte contre cette maladie. Des études ont par ailleurs mis en évidence un certain nombre de QTL associés à la résistance à cette maladie (Vallejo et al. (1998), Yonash et al. (1999), Cheng (2005)). D'autres marqueurs pour la résistance aux maladies ont été mis en évidence chez les bovins avec la dermatophilose (Maillard et al. (2003)), la diarrhée causée par le *E. coli* chez les porcs (Edfors & Wallgren (2000)) et la tremblante chez les moutons (Hunter et al. (1996)).

1.4 Le programme français d'éradication de la tremblante

La tremblante n'est pas considérée comme une zoonose. Le ministère de l'agriculture s'est intéressé à cette maladie à cause de la relation entre la MCJ et l'ESB. Même si aucun lien direct n'a été mis en évidence entre la Tremblante et la MCJ, la transmission de l'ESB aux petits ruminants a été cependant démontrée. C'est pourquoi, la possibilité de transmission ne peut être exclue. Un programme d'éradication a été mis en place dans le but de garantir la sécurité du consommateur et la santé animale.

Une étude a montré que les ovins génétiquement résistants à la tremblante classique sont résistants à l'ESB (Jeffrey et al. (2001)). L'utilisation de mâles ARR/ARR comme reproducteurs semble donc être un bon moyen pour réduire la contamination en élevage. De plus, l'absence de lien entre les caractères de reproduction, de production et le génotype PrP a été démontrée (Palhière et al. (2002), Brochard & Palhière (2004), Vitezica et al. (2006), Brochard et al. (2006)). Toutes ces informations permettent donc d'envisager l'utilisation du polymorphisme du gène PrP afin d'améliorer la résistance à la tremblante.

Le dispositif du ministère de l'Agriculture français pour lutter contre la tremblante comprend différents éléments complémentaires : (source : <http://agriculture.gouv.fr/sections/thematiques/sante-protection-animaux/maladies-animales/tremblante>)

- Un réseau de surveillance clinique. Des tests biochimiques « rapides » de dépistage sont réalisés sur 40 000 ovins et 20 000 caprins à l'abattoir (destinés à la consommation) et sur 20 000 ovins et 17 000 caprins à l'équarrissage (morts ou malades). Les analyses sont réalisées sur le tronc cérébral des animaux. Les carcasses détectées positives à l'abattoir sont saisies et détruites. Des mesures de police sanitaire sont mises en place dans le troupeau d'origine de l'animal atteint, si l'identification et la traçabilité le permettent.
- Des mesures de police sanitaire sont mises en oeuvre dès la détection d'un cas, un arrêté préfectoral portant déclaration d'infection est pris. Tous les ovins du ou des élevages concernés sont recensés et génotypés afin de déterminer leur degré de résistance à la tremblante. Les ovins porteurs du génotype sensible et très sensible sont euthanasiés et détruits. Les ovins « résistants » sont conservés. Les locaux sont désinfectés puis l'arrêté préfectoral est levé. Ensuite des ovins résistants sont utilisés pour le repeuplement des troupeaux. Dans les élevages détenteurs d'ovins présentant une sensibilité particulière à la tremblante (plus de 20% d'animaux sensibles et très sensibles), un programme d'amélioration génétique est mis en place en accord avec l'éleveur. Cet éleveur pourra conserver pendant une période maximale de 2 ans ces femelles reproductrices « sensibles » et les faire se reproduire avec des mâles résistants. Les agnelles de renouvellement seront donc résistantes. Ces mesures permettent d'éviter l'abattage total des troupeaux peuplés d'animaux de race sensible. A la fin de cette période, les femelles sensibles seront euthanasiées.
- Un plan d'amélioration génétique basé sur l'utilisation du polymorphisme du gène PrP comprenant quatre objectifs :
 - Elimination des animaux porteurs de l'allèle VRQ.
 - Aide aux éleveurs dont les troupeaux sont atteints. L'union nationale de sélection et de promotion de race (UPRA) pourrait contribuer à répondre aux besoins des éleveurs des troupeaux atteints.
 - Sélection des animaux résistants dans le noyau de sélection. L'objectif est que

100% des géniteurs mâles et femelles soient ARR/ARR.

- Utilisation des béliers ARR/ARR pour la production d'agneaux de boucherie. La commercialisation d'agneaux porteurs de l'allèle ARR afin de protéger les consommateurs d'un risque éventuel de transmission d'une ESST (notamment l'ESB). Pour cela, l'identification et la reproduction des béliers ARR/ARR doivent se faire en grande quantité.

En 2008, l'allèle VRQ n'est plus présent chez les mâles reproducteurs des entreprises de sélection. Concernant la voie femelle, l'allèle VRQ est quasiment éliminé. Chez les ovins allaitants, 93% des béliers sont résistants (Tableau 1.5) alors qu'en 2002 ils ne représentaient que 24% (Tableau 1.4). Concernant les ovins lait, 85% des mâles pour l'insémination artificielle (principal mode de reproduction) et 95% des doses d'insémination artificielle sont de génotype ARR/ARR contre 31% en 2002. Les objectifs du programme d'amélioration génétique pour la résistance à la tremblante ont été atteints.

1.5 Conclusion

La contribution de la génétique aux stratégies de lutte contre les maladies apparaît comme indispensable. En effet, la durabilité et l'efficacité de nombreuses autres méthodes déjà mises en place pourraient être améliorées si une telle sélection génétique est mise en place. De nombreuses études sur la variation intra et interrassiale en matière de sensibilité à de nombreuses maladies ont été réalisées et, dans un certain nombre de cas, les résultats issus de ces études ont été inclus dans des programmes de sélection. La sélection assistée par gène ou par marqueurs a de nombreux avantages :

- permanence des effets une fois établis ;
- diminution des coûts pour combattre ces maladies (vétérinaires, traitements chimiques, etc) ;
- efficacité prolongée des méthodes de contrôles ;
- possibilité d'effets plus larges (accroître la résistance à plus d'une maladie).

La sélection assistée par gène ou par marqueurs apparaît donc comme une solution complémentaire satisfaisante pour combattre ces maladies.

TABLE 1.5 – Fréquences des allèles du gène PrP dans les troupeaux ovins français (agneaux et agnelles) en 2008

Race	ARR	AHQ	ARQ	VRQ
Ile de France	0.964	0.000	0.030	0.006
Rava	0.748	0.001	0.249	0.001
Charmoise	0.814	0.003	0.169	0.015
Berrichon du Cher	0.954	0.004	0.042	0.000
Est à laine Merinos	0.477	0.044	0.478	0.001
Limousine	0.748	0.000	0.250	0.002
Bleu du maine	0.969	0.000	0.014	0.017
Texel	0.864	0.011	0.123	0.002
Suffolk	0.925	0.001	0.074	0.000
Hampshire	0.782	0.000	0.215	0.003
Merinos d'Arles	0.736	0.012	0.250	0.002
Préalpes du sud	0.785	0.003	0.212	0.000
Bizet	0.808	0.000	0.192	0.000
Noir du Velay	0.705	0.000	0.289	0.007
Lacaune viande	0.875	0.002	0.122	0.000
Blanc du massif central	0.712	0.015	0.266	0.007
Causses du Lot	0.748	0.043	0.196	0.013
Grivette	0.664	0.003	0.334	0.000
Moutons vendéens	0.766	0.000	0.233	0.001
Rouge de l'Ouest	0.917	0.000	0.080	0.003
Tarasconnaise	0.712	0.009	0.278	0.001
Lacaune lait	0.939	0.001	0.058	0.002
Basco-Béarnaise	0.577	0.001	0.421	0.000
Corse	0.547	0.074	0.379	0.000
Manech tête noire	0.607	0.010	0.383	0.000
Manech tête rousse	0.663	0.001	0.336	0.001

Chapitre 2

Modélisation de la sélection

2.1 Introduction

La sélection des animaux domestiques permet d'augmenter les performances moyennes des différentes populations sur le plan économique (quantité de lait ou de viande, qualité des produits) de manière pérenne. Avant de mettre en place un programme de sélection, il est important de définir clairement les objectifs et les critères de sélection. Un programme de sélection combine évaluation, sélection et accouplement des reproducteurs sur plusieurs générations. L'évaluation permet de connaître les performances des différents individus de la population afin de sélectionner les « meilleurs » animaux selon les critères fixés préalablement. Ces derniers seront accouplés pour obtenir une nouvelle génération avec des performances en moyenne plus élevées.

2.2 Objectif et critère de sélection

La définition des objectifs de sélection constituent la première étape dans l'organisation d'un programme de sélection. Ils correspondent à un ensemble de caractères que l'on souhaite améliorer. Ces caractères sont parfois antagonistes et/ou non mesurables sur l'individu lui-même (qualité et/ou quantité de lait, qualité et/ou quantité de viande). Le choix des objectifs dépend des conditions d'élevage, du marché des produits, des attentes des consommateurs. De ce fait différents acteurs interviennent dans le choix de ces objectifs. Les objectifs dans de nombreux programmes de sélection correspondent à l'amélioration de la qualité du produit et à la diminution des coûts de production. On définit un objectif global G comme une combinaison linéaire des caractères affectés d'un coefficient de pondération.

Il convient de distinguer les caractères qui définissent l'objectif de sélection et ceux qui constituent les critères de sélection. Les premiers peuvent ne pas être mesurables en pratique sur les candidats à la sélection, mais sont en corrélation avec les seconds. Ces critères correspondent donc à un caractère ou à une combinaison de caractères que l'on peut mesurer ou calculer à partir de mesures effectuées sur les candidats à la sélection et/ou sur leur apparentés. Ces critères font généralement l'objet de calcul d'indice de sélection I permettant de classer les animaux reproducteurs. Ces critères doivent être, dans l'idéal, précoces pour limiter l'intervalle de génération et de mesures peu coûteuses. Ils doivent avoir une héritabilité suffisante afin que l'impact de la sélection se fasse ressentir sur les générations futures. Ces

critères de sélection permettent de classer les individus et de choisir les meilleurs pour la reproduction. Ce classement se fait sur une estimation de leur valeur génétique G .

2.3 Indexation

2.3.1 Principes généraux de l'indexation

Le calcul de l'indice I de chaque individu utilise les mesures pouvant provenir de l'individu lui-même et/ou de ses apparentés. Plus la quantité d'information est abondante et informative (prise sur des apparentés proches de l'individu), plus l'estimation de G par I sera précise. Différentes méthodes d'indexation sont possibles dépendant de la provenance des informations : si l'information provient de l'individu lui-même on parle d'évaluation sur performance propre ; si elle provient des parents ou grands parents, on parle d'évaluation sur ascendance ; si les informations proviennent de ses enfants, on parle d'évaluation sur descendance. Ces diverses méthodes d'indexation sont le plus souvent combinées. En pratique, la valeur génétique G d'un animal est évaluée par un indice I prenant en compte l'ensemble des relations de parentés entre tous les individus mesurés et tous les candidats à la sélection (modèle « animal »).

2.3.2 Différentes méthodes d'indexation

Sélection massale

Les animaux sont évalués sur leurs propres performances (le poids ou la prolificité). Cette méthode est la plus simple de mise en oeuvre et très avantageuse pour les caractères à forte héritabilité. Néanmoins, elle n'est pas applicable quand le caractère est mesuré après l'abattage de l'animal (exemple : qualité de viande, mesure sur la carcasse) ou n'est pas exprimé chez le candidat à la sélection (production laitière pour un taureau). Pour pallier à certaines de ces limites, des mesures de caractères très fortement corrélés avec le caractère initial peuvent être effectuées. Cependant, d'autres méthodes d'évaluation prenant en compte la généalogie de l'animal (ascendants, descendants) s'avèrent plus adaptées à tous les caractères.

Évaluation sur ascendance

Cette méthode d'évaluation permet de choisir précocement les futurs reproducteurs ce qui constitue son avantage majeur dans la mise en place d'un plan de sélection. En effet, l'évaluation de ces derniers peut être faite avant leur naissance. Néanmoins, pour mettre en place cette méthode d'évaluation, il est nécessaire d'avoir des fichiers où la généalogie ainsi que les performances soient répertoriées de manière régulière et fiable ce qui peut être une difficulté dans certains pays en développement. Son gros inconvénient se trouve dans sa faible précision. C'est pourquoi, le choix des reproducteurs doit être affiné dans un second temps avec d'autres méthodes d'évaluation comme la méthode d'évaluation individuelle ou sur descendance.

Evaluation sur descendance

Cette méthode donne une grande précision dans l'estimation de la valeur génétique et s'applique à tous les caractères. Cette précision est très importante lorsque les animaux sont destinés à être diffusés de manière conséquente ou lorsque des accouplements raisonnés sont envisagés. Néanmoins, cette évaluation est tardive puisqu'il faut attendre que les descendants soient mesurés. Le risque de perdre le candidat (accident, maladie) est non négligeable. En outre, le fait que cette évaluation soit tardive entraîne un coût non négligeable pour les éleveurs, car les animaux sont entretenus jusqu'à l'obtention de leur évaluation, certains n'étant finalement pas conservés au vu des résultats. Le coût élevé et l'organisation rigoureuse de cette évaluation incite à n'avoir qu'un nombre réduit de candidats qui sont triés préalablement par d'autres méthodes d'évaluation (massale, sur ascendance).

2.4 Réponse à la sélection et progrès génétique

Le but de la sélection est d'améliorer de génération en génération la valeur génétique moyenne de la population. Parmi les candidats à la sélection de la génération t , les reproducteurs sont sélectionnés et vont produire la génération $t + 1$. Le résultat de la sélection est un progrès génétique ΔG réalisé, différence entre la valeur génétique moyenne de la génération t et celle de la génération $t + 1$. Comme l'intervalle de génération peut varier d'une espèce à l'autre ou d'une méthode de sélection à l'autre, il est préférable de considérer le progrès génétique exprimé par unité de temps ΔG_a :

$$\Delta G_a = \frac{\Delta G}{\text{Intervalle de génération}}$$

La sélection ne se fait généralement pas de la même manière dans les deux sexes, avec des différences en terme de pression de sélection et de méthode d'évaluation. Dans la transmission des gènes entre générations, on doit distinguer quatre voies, car un individu peut être père (P) ou mère (M) d'un futur père ou mère. La formule du progrès génétique donnée par Rendel et Robertson en 1950 est :

$$\Delta G = \frac{i_{PP}r_{PP} + i_{PM}r_{PM} + i_{MP}r_{MP} + i_{MM}r_{MM}}{t_{PP} + t_{PM} + t_{MP} + t_{MM}} \sigma_a$$

où i représente l'intensité de sélection, r la précision de la méthode d'évaluation, t l'intervalle de génération et σ_a est la racine carré de la variance génétique additive du caractère soumis à sélection. Le progrès génétique dépend donc de 4 paramètres qui vont être détaillés par la suite :

- intervalle de génération ;
- variabilité génétique ;
- intensité de sélection ;
- précision de la sélection.

2.4.1 Intervalle de génération

L'intervalle de génération est l'âge moyen des parents à la naissance de leurs descendants (Haldane (1932), Lush (1937)) susceptibles d'être conservés pour la reproduction (Figure 2.1). Le progrès génétique obtenu est d'autant plus grand que l'intervalle de génération est court.

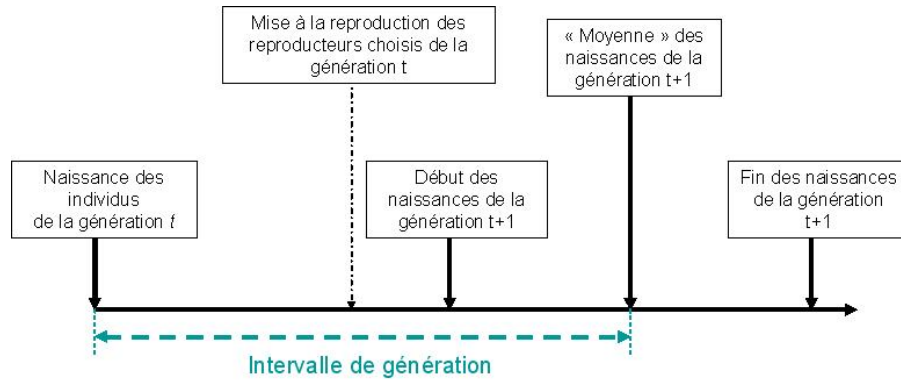


FIGURE 2.1 – Intervalle de génération entre 2 générations successives t et t+1

2.4.2 Variabilité génétique

La variabilité génétique est la source biologique indispensable à tout progrès génétique, qui est d'autant plus grande, à même modalité de sélection, que cette variabilité est plus importante. Néanmoins, la taille réduite des populations et l'utilisation importante d'un faible nombre de reproducteurs augmentent la consanguinité et diminuent la variabilité génétique. Pour garder les ressources nécessaires en cas de changement d'objectifs de sélection dans les populations d'animaux d'élevage, il est donc nécessaire de maintenir dans la population une certaine diversité génétique par un choix judicieux des reproducteurs et de leurs accouplements, voire par l'introduction de reproducteurs étrangers à la population.

2.4.3 Intensité de sélection

L'intensité de sélection i exprime la force de la sélection. Elle varie à l'inverse de la proportion d'animaux sélectionnés. L'intensité de sélection mesure la différence entre la moyenne des reproducteurs sélectionnés et la moyenne des candidats à la sélection. Soit un critère de sélection distribué selon une loi normale de moyenne μ_0 et de variance σ^2 (Figure 2.2). Les animaux ayant une performance supérieure au seuil de sélection fixé (soit une proportion q des candidats) sont sélectionnés. Leur moyenne est de μ_s . La différence S , entre les deux moyennes ($\mu_s - \mu_0$) appelée différentielle de sélection se calcule dans le cas d'une loi normale comme $S = i(q)\sigma_a = \frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\lambda^2}}{q}\sigma_a$

2.4.4 Précision de la sélection

La précision de la sélection notée $r(G, I)$ est mesurée par la corrélation entre la valeur génétique vraie G et son estimation I .

2.5 Schéma de sélection

Le schéma de sélection est une combinaison de plusieurs choix :

- méthode d'évaluation,
- intervalle de génération,
- intensité de sélection,

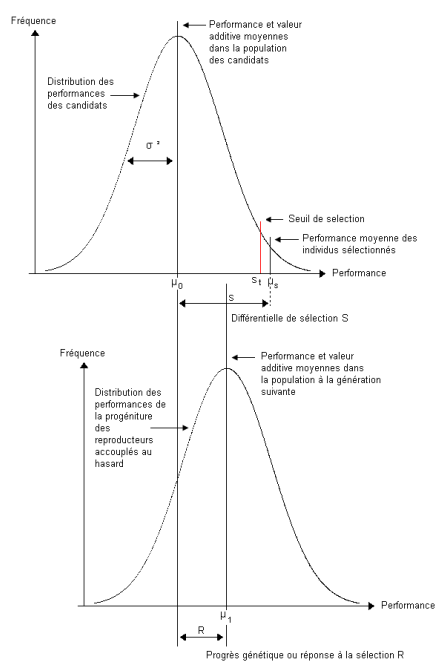


FIGURE 2.2 – Une génération de sélection (Minvielle (1998))

- mode d'accouplements.

Cette combinaison a pour objectif d'obtenir le plus grand progrès génétique possible, année après année. Ces choix sont opérés compte tenu des spécificités de l'espèce animale et des caractères étudiés. De plus, il est nécessaire de connaître les liens entre les différents facteurs du progrès génétique :

Relation entre l'intensité de sélection i et l'intervalle de génération t

L'intervalle de génération dépend du temps d'utilisation des reproducteurs. L'augmentation de ce temps d'utilisation implique un intervalle de génération plus long et une augmentation de l'intensité de sélection. En effet, pour un nombre de reproducteurs constant, le besoin en reproducteurs pour le remplacement des parents diminue.

Relation entre l'intensité de sélection i et la précision de sélection r

La diminution du taux de sélection augmente l'intensité de sélection. Le nombre de reproducteurs est en général fixé, il faut donc que le nombre de candidats soit plus grand. Dans le cas de la sélection sur descendance, en augmentant le nombre de candidats à la sélection, on diminue le nombre de descendants contrôlés par candidat à taille constante de la population mesurable. Par conséquent la précision de la sélection (fonction du nombre de descendants contrôlés) décroît.

Relation entre l'intervalle de génération t et la précision de sélection r

Afin d'augmenter la précision de sélection, on peut avoir recours à la sélection sur descendance ou augmenter le nombre de performances contrôlées en sélection individuelle à condition que le caractère soit répétable. Dans ces conditions l'intervalle de génération augmente.

La modélisation d'un schéma de sélection permet de prévoir l'évolution de la population soumise à ce schéma. Elle va décrire la structure générale de la population ainsi que son organisation à l'aide d'équations reliant différentes variables aboutissant à une fonction objective. Pour caractériser la population, celle-ci sera subdivisée en classes (par exemple sexe, âge catégorie,...). Ces classes diffèrent d'un schéma de sélection à l'autre. Les individus d'une même classe sont soumis aux mêmes règles de décision (sélection, accouplement). A chaque classe sont associées des variables d'état qui caractérisent la population du point de vue génétique (par exemple l'espérance de la valeur génétique pour un caractère polygénique des individus de la classe, leur fréquence génotypique pour un gène majeur). Les valeurs prises par ces variables évoluent pendant le processus de sélection. La modélisation doit décrire cette évolution. D'autres variables interviennent dans la description du processus de sélection : les paramètres, les variables de décision et des variables internes. Les paramètres sont déterminés par la « nature », ils peuvent être démographiques (taille de la population, taux de mortalité...), économiques (coût engendré par les mesures, coût d'entretien...), génétiques (héritabilité du caractère) et zootechniques (pourcentage d'IA). Les variables de décision permettent de décrire les composantes d'un schéma de sélection sur lesquels l'opérateur peut agir (taux de sélection, proportion de mâles élites, durée d'utilisation des mâles, nombre de descendants contrôlés par mâles évalués...). Les variables internes permettent de faire le lien entre les paramètres, les variables de décision et les variables d'état (intervalle de génération, corrélation entre la valeur génétique et son estimation,...). Lors

de la mise en place d'un schéma de sélection, certaines contraintes apparaissent et sont en général des contraintes démographiques, biologiques ou organisationnelles apparaissant sous forme d'équations ou d'inéquations. Par exemple, les taux de sélection sont des proportions, donc compris entre 0 et 1, le nombre de mâles en testage est limité ou les mâles testés sont des fils de mâles élités.

2.6 Sélection assistée par marqueurs

L'utilisation des marqueurs moléculaires (caractère génétique très facile à observer en laboratoire) est de plus en plus répandue. Les marqueurs moléculaires sont des fragments d'ADN. Ils sont utilisés dans divers domaines (analyse des caractères quantitatifs, évaluation de la biodiversité, paléogénétique, enquêtes policières, recherche en paternité, etc...). Ils permettent de suivre la transmission des caractéristiques génétiques entre générations. Par exemple si un individu est porteur d'un allèle a , que son père possède cet allèle et que sa mère ne le possède pas, on en déduit que l'individu l'a hérité de son père. Ces marqueurs sont des indicateurs de la variabilité génétique permettant d'identifier le polymorphisme entre espèces, races, individus. Lors de la fabrication des gamètes, les allèles en un QTL et un marqueur proche sont le plus souvent cotransmis. Ce phénomène est utilisé pour la détection des QTL et permet de sélectionner les individus porteurs du marqueur indiquant la présence de l'allèle favorable au QTL. Cette nouvelle sélection est appelée la sélection assistée par marqueurs (SAM). Son principe est décrit dans le schéma suivant (Figure 2.3) :

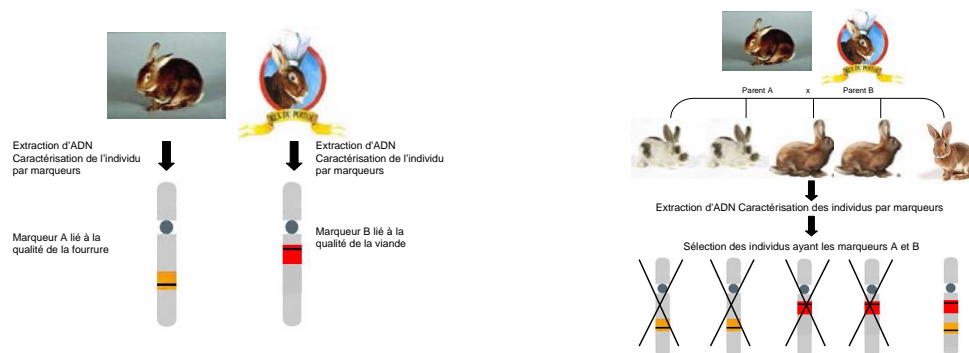


FIGURE 2.3 – Principe de la sélection assistée par marqueurs

L'efficacité de la SAM comparée à la sélection classique dépend principalement du caractère étudié et de la durée de la sélection. La SAM permet d'obtenir de meilleurs résultats dans le cas où la sélection classique est peu efficace et les informations moléculaires très détaillées. Cette sélection est donc souvent utilisée avec des caractères peu héritables et/ou avec des caractères exprimés dans un seul sexe. Avec une approche déterministe, Ollivier (Ollivier (1998)) a montré que l'utilisation des marqueurs dans des index combinés n'améliore que légèrement la précision de la sélection. Avec une approche stochastique, Ruane et Colleau (Ruane & Colleau (1995)) ont trouvé que l'apport de la SAM, pour des caractères d'héritabilité moyenne, était modeste. De plus, ils ont montré que l'évolution des moyennes polygéniques était défavorable par rapport à l'indexation BLUP basée sur les performances.

Néanmoins, Ruane et Colleau en 1996 (Ruane & Colleau (1995)) ont montré qu'une SAM dans laquelle est introduite une sélection génotypique de jeunes mâles pour l'amélioration de caractères exprimés dans un seul sexe s'avère plus efficace que la sélection classique pour les gains polygéniques et permet l'augmentation de la fréquence des allèles favorables au QTL.

De nombreuses études ont montré qu'à court terme, la SAM améliorait les résultats relatifs au progrès génétique (Smith (1967), Lande & Thompson (1990), Meuwissen & Goddard (1996), Van der Beek et al. (1995), Ruane & Colleau (1995), Ruane & Colleau (1996), Spelman & Van Arendonk (1997)). Sur le long terme, Gibson (Gibson (1994)) a montré, par des simulations multi générations, que la SAM donnait des gains génétiques inférieurs et des taux de consanguinité supérieurs à ceux produits par la sélection sur performances. De même, l'étude de Verrier (Verrier (2001)) montre que les allèles favorables ont été fixés plus rapidement qu'en sélection classique mais que la réponse polygénique est plus faible. Par conséquent, sur le long terme, la SAM apparaît moins performante que la sélection classique.

Néanmoins Dekkers et Van Arendonk (Dekkers & Van Arendonk (1998)) ainsi que Manfredi et al (Manfredi et al. (1998)) ont montré que les résultats obtenus étaient supérieurs à ceux de la sélection classique si on construit une SAM dynamique, c'est à dire que les poids relatifs donnés au QTL et aux polygènes dans l'index combiné évoluent dans le temps. Ces résultats étaient supérieurs en terme de vitesse de fixation des allèles favorables au QTL, d'évolution des moyennes polygéniques et des gains génétiques totaux pour tout horizon donné.

L'utilisation des typages en dehors des index combinés offre des valeurs ajoutées pour la sélection. En effet, elle permet de diminuer le nombre de générations pour atteindre le ou les objectifs (Hospital et al. (1992), Visscher et al. (1996), Hospital & Charcosset (1997)). Des phases de sélection génotypique peuvent s'insérer dans un schéma de sélection massale. Cependant, les cas souvent étudiés correspondent à des schémas plus complexes. Différentes conditions d'utilisation des typages peuvent être envisagées :

- une présélection génotypique de jeunes mâles avant leur mise en testage sur descendance. Les avantages de cette présélection, cités précédemment, ont été montré par Dekkers et Van Arendonk (Dekkers & Van Arendonk (1998)) ainsi que Manfredi et al (Manfredi et al. (1998)),
- un index combiné pour la sélection des adultes,
- des accouplements raisonnés entre parents, c'est-à-dire des accouplements en fonction des génotypes au gène/QTL considéré.

Avant toute application de la SAM, une estimation financière est nécessaire car les procédés permettant d'associer un allèle marqueur à un allèle du QTL pour chaque reproducteur peuvent devenir coûteux dans des programmes de sélection en routine (Soller & Medjugorac (1999)). Par conséquent, la SAM doit être considérée comme un outil complémentaire aux méthodes d'indexation et aux schémas de sélection classiques.

2.7 Optimisation de la sélection

Grâce à la modélisation du programme de sélection, une fonction objective peut être évaluée et optimisée. Cette fonction peut être de différentes natures : technique (progrès génétique annuel ou final, taux de consanguinité, variabilité génétique) ou économique (coûts occasionnés par le contrôle de performances, par les inséminations artificielles, recettes de la production). L'optimisation de cette fonction objective correspond à une maximisation/minimisation d'une fonction mathématique sous contraintes.

Afin d'exploiter les différentes potentialités de la SAM (Colleau (1999), Fritz et al. (2003)), de nombreuses études ont été faites sur l'optimisation de la sélection assistée par marqueurs : nouvelles étapes de sélection, accouplement raisonnés, fixation d'un allèle favorable. L'une des grandes préoccupations suscitée par la SAM est l'augmentation de la consanguinité donc une perte de la variabilité génétique. Des méthodes ont été développées pour gérer l'évolution de cette consanguinité (Wray & Goddard (1994), Grundy et al. (1998)), en terme de maximisation du progrès génétique tout en contraignant le taux de consanguinité (1% par génération en général) ou de minimisation de la consanguinité tout en effectuant un certain progrès génétique (Colleau et al. (2004)).

Concrètement le but de ces études est d'optimiser la sélection des reproducteurs et leur contribution (Colleau et al. (2004), Pong-Wong & Woolliams (2007), Villanueva et al. (2002)). D'autres études ont fait la comparaison entre l'optimisation de la sélection assistée par marqueur et l'optimisation de la sélection classique (Villanueva et al. (2004), Woolliams et al. (2002)). Diverses méthodes existent pour résoudre ce genre de problème. Les principales méthodes d'optimisation utilisées sont :

- méthode de Lagrange (Meuwissen (1997), Hinrichs et al. (2006)),
- méthode de Lagrange et recuit simulé (Kirkpatrick et al. (1983), Sonesson & Meuwissen (2000), Sonesson & Meuwissen (2002), Colleau & Tribout (2008)),
- *semidefinite programming* (SDP) (Vandenberghe et al. (2003), Pong-Wong & Woolliams (2007))
- programmation dynamique (Lewis (1986), Dekkers & Van Arendonk (1998), Villanueva et al. (2004)),
- *sequencial quadratic programming* (SQP) (Manfredi et al. (1998)).

Toutes ces méthodes d'optimisation vont être décrites succinctement dans la partie suivante et une autre méthode d'optimisation, méthode utilisée lors de ce travail, sera décrite de manière plus détaillée : les algorithmes génétiques.

Chapitre 3

Optimisation

3.1 Modèle

Un modèle mathématique décrit un phénomène réel. Il correspond à un ensemble de règles ou d'équations composée de N variables décisionnelles X (variables détaillant la situation que l'on souhaite étudier) et va permettre de répondre à une ou des questions que l'on se pose : les objectif à atteindre. Ces objectifs sont transcrits en une fonction mathématique composée des variables décisionnelles, appelée fonction objective f . Les variables décisionnelles sont, en général, limitées par la description que l'on en fait. Ces limites sont décrites par des paramètres B . L'ensemble de ces paramètres limitant le modèle est associé aux contraintes G . Ces contraintes sont des équations ou inéquations fonction des variables de décision et des paramètres. Après la modélisation du phénomène à étudier, l'optimisation du modèle peut être envisagée. Rappelons que l'objectif en optimisation est de choisir parmi l'ensemble des configurations possibles des différentes variables, la configuration qui satisfasse les contraintes et donne une solution optimale (maximale ou minimale) pour la fonction objective f . Ce problème s'écrit sous la forme :

$$\underset{X \in \mathbb{R}^N}{Max} f(X) \text{ ou } \underset{X \in \mathbb{R}^N}{Min} f(X)$$

$$\text{Sous } G(X) \leq B$$

La stratégie de résolution d'énumérer toutes les solutions possibles et de conserver la meilleure est utopique. En effet, le nombre de solutions croît de manière exponentielle avec la taille du problème.

On peut regrouper les différentes méthodes d'optimisation en deux familles. La première est la famille des méthodes dites « *exactes* ». Dans le cas où la fonction objectif f possède une dérivée seconde continue, les problèmes sans contraintes peuvent être résolus en trouvant les points X^* qui annulent le gradient Δf et en calculant la valeur de la hessienne (matrice carrée des dérivées partielles secondes) $\Delta^2 f$ aux points X^* . Si cette valeur est positive, le point obtenu est un *minimum local*, si elle est négative c'est un *maximum local*. Si la fonction objectif est convexe, le minimum (resp. maximum) est un minimum (resp. maximum) global. Dans le cas de problèmes avec contraintes d'égalité, on peut se ramener à un problème sans contraintes en utilisant les multiplicateurs de Lagrange à condition que la fonction soit dérivable.

L'utilisation de la deuxième famille de méthodes d'optimisation est en plein essor depuis les années 1980 alors que les premiers algorithmes datent de 1951 (Robbins et al.). Ces méthodes sont appelées « les métaheuristiques » et sont utilisées pour résoudre des problèmes d'optimisation difficile pour lesquels on ne connaît pas de méthode classique plus efficace. Ces algorithmes ont une part d'aléatoire, ce qui permet une meilleure exploration de l'espace des solutions. Les métaheuristiques s'inspirent en général des systèmes naturels en biologie avec l'algorithme génétique (Holland (1962a,b)), l'algorithme de colonies de fourmis (Dorigo & Gambardella (1997)), les réseaux de neurones (Hopfield (1982)), en physique avec le recuit simulé (*simulated annealing*, Kirkpatrick et al. (1983)). En 1987, David Ackley (Ackley (1987)) a comparé différents métaheuristiques. Aucun de ces algorithmes n'émerge en tant que « vainqueur ». Néanmoins, quand la fonction objectif devient plus complexe, les algorithmes génétiques et le recuit simulé apparaissent comme plus performants.

Ces méthodes présentent chacune des avantages et des inconvénients dans la résolution des problèmes d'optimisation globale. Afin d'exploiter la potentialité de chacune de ces méthodes, de nouvelles méthodes couplées dites « hybrides » apparaissent depuis une quinzaine d'années. Elles ont pour but d'améliorer les performances des algorithmes d'optimisation telle que le temps d'exécution, le critère de convergence et la précision de la solution obtenue.

3.2 Méthodes d'optimisation

3.2.1 Méthode de Lagrange

Le problème à résoudre se formule de la manière suivante :

$$\min_{x \in G} f(x) \quad \text{avec} \quad G = \{ x \in E, g(x) = 0 \}$$

Ce problème peut être résolu par l'introduction des multiplicateurs de Lagrange. Le problème s'écrit alors sous la forme :

$$\forall (x, \lambda) \in E \times F \quad \min_{(x, \lambda) \in E \times F} L(x, \lambda) = f(x) - \lambda g(x)$$

Pour résoudre ce nouveau problème sans contraintes, les dérivées partielles sont calculées et le système (S) suivant est résolu afin d'obtenir les valeurs de x et de λ :

$$(S) \quad \begin{cases} \frac{\partial L(x, \lambda)}{\partial x} = 0 \\ \frac{\partial L(x, \lambda)}{\partial \lambda} = 0 \end{cases}$$

Cette méthode peut être généralisée aux problèmes d'optimisation incluant des contraintes d'inégalités (ou non linéaires) ou sur des fonctions convexes en utilisant la méthode des multiplicateurs d'Everett (ou multiplicateurs généralisés de Lagrange), plus fréquemment appelée méthode des pénalités (Everett (1963)).

3.2.2 Sequential Quadratic Programming SQP

La méthode SQP est une méthode d'optimisation pour des problèmes avec des contraintes de nature différentes :

- contraintes d'inégalités non linéaires,
- ou contraintes d'inégalités linéaires,
- ou contraintes d'égalités non linéaires.

Le problème s'écrit sous cette forme :

$$\min_{x \in X} f(x)$$

sous

$$g_i(x) = 0 \quad i = 1, \dots, n_1$$

$$g_j(x) \geq 0 \quad j = n_1 + 1, \dots, n_2$$

Les fonctions f et g doivent être continues et deux fois dérivables. Le principe de cette méthode consiste à reformuler ce problème en un problème de programmation quadratique par une approximation quadratique du Lagrangien de la fonction objectif et une linéarisation des contraintes. On effectue une approximation de Taylor à l'ordre 2 de la fonction objective à l'itération k , :

$$f(x_k + \delta x_k) \approx f(x_k) + \nabla f(x_k) \delta x_k + \frac{1}{2} \delta x_k^T \nabla^2 f(x_k) \delta x_k$$

Puis la méthode de Lagrange est utilisée afin d'obtenir le sous problème suivant :

$$\min_{x_k, \lambda_k, \beta_k} L(x_k) + \nabla^T L(x_k) \delta x_k + \frac{1}{2} \delta x_k^T \nabla^2 L(x_k, \lambda_k, \beta_k) \delta x_k$$

sous

$$g_i(x_k) + \nabla^T g_i(x_k) \delta x_k = 0 \quad i = 1, \dots, n_1$$

$$g_j(x_k) + \nabla^T g_j(x_k) \delta x_k \geq 0 \quad j = n_1 + 1, \dots, n_2$$

où $\delta x_k = x - x_k$ et L fonction de Lagrange s'écrivant sous la forme $L(x_k, \lambda_k, \beta_k) = f(x_k) + \lambda_k g_i(x_k) + \beta_k g_j(x_k)$ où λ et β sont les multiplicateurs de Lagrange.

L'algorithme de cette méthode est le suivant :

1. Initialisation du problème. Matrice Hessienne initialisée par la matrice identité
2. Evaluation de la fonction objectif et des contraintes
3. Evaluation des gradients de la fonction objectif et des contraintes
4. Résolution de sous problème quadratique
5. Vérifications des conditions d'arrêt si vérifiées solution trouvée sinon :
 - (a) Estimation de la matrice Hessienne
 - (b) Estimation des paramètres de Lagrange et Kuhn-Tucker
 - (c) Estimation de la direction de descente δx_k
 - (d) Estimation de la longueur du pas α_k
 - (e) Calcul du nouveau point $x_{k+1} = x_k + \alpha_k \delta x_k$

6. Retour à l'évaluation de la fonction objectif et des contraintes (2)

Contrairement à d'autres méthodes d'optimisation, la méthode SQP ne vérifie pas à chaque itération si toutes les contraintes sont respectées mais impose simplement qu'aucune contrainte ne soit violée pour la solution finale. Un point important de cette méthode est l'estimation des gradients. Il est nécessaire d'avoir une bonne précision sur l'estimation des gradients puisqu'ils déterminent la direction de la descente et interviennent dans les conditions d'arrêt de l'algorithme.

3.2.3 Programmation semi-définie SDP

Un problème de programmation semi-définie est un problème qui est de la forme :

$$\min f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n a_{ij} x_i x_j + \sum_{i=1}^n b_i x_i$$

sous

$$g_j(x_1, x_2, \dots, x_n) = \sum_{i=1}^n c_{i,j} x_i \leq 0 \quad \forall j \in 1..m$$

avec a_{ij} éléments d'une matrice A semi-définie positive c'est à dire $\forall X \in \mathbb{R}^n : X^T A X \geq 0$. Le problème de programmation semi-définie est un cas particulier des problèmes de programmation quadratique. Pour résoudre ce problème avec contraintes, on se ramène à un problème sans contrainte en utilisant la fonction de Lagrange L :

$$L(x, \lambda) = f(x) + \sum_{j=1}^m \lambda_j g_j(x)$$

La fonction L est une fonction de la variable primale x et de la variable duale λ . Les contraintes $g_j \leq 0$ se traduisent par des contraintes sur la variable $\lambda : \lambda_j \geq 0$. On définit la fonction duale Θ de la manière suivante :

$$\Theta(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda)$$

Le problème d'optimisation à résoudre devient alors la résolution du problème dual :

$$\max \Theta(\lambda) \quad \text{avec} \quad \lambda \in \mathbb{R}^{+m}$$

La méthode de points intérieurs est utilisée pour résoudre ce genre de problème. Cet algorithme a été introduit par Karmarkar en 1984 (Karmarkar (1984)) pour la programmation linéaire et a été étendu dans le cadre de l'optimisation convexe (Nesterov & Nemirovski (1994)), et donc pour la programmation semi-définie.

3.2.4 Programmation dynamique

La programmation dynamique permet de résoudre des problèmes décisionnels évoluant sur plusieurs étapes ($t \in [0, T]$). La transition de l'étape t à $t+1$ est décrite par des équations différentielles du premier ordre dépendant de variables d'états $x(t)$ et de variables de décisions $u(t)$. Le problème s'écrit alors de la forme suivante :

$$\dot{x}(t) = f(x(t), u(t))$$

De là, une fonctionnelle de gain J est définie en prenant en compte toutes les étapes et un gain final h . Cette fonctionnelle s'écrit alors sous la forme suivante :

$$J(u) = \sum_{t=1}^{T-1} g(x(t), u(t), t) + h(x(T))$$

Les fonctions g et h sont respectivement appelées fonction de gain instantané et fonction de gain final. Par conséquent le problème initial à résoudre devient :

$$\min J(u) = \min \sum_{t=1}^{T-1} g(x(t), u(t), t) + h(x(T))$$

Le principe de Bellman est appliqué. Ce principe dit que les parties d'une trajectoire optimale sont elles-mêmes optimales. Ce principe permet d'obtenir l'équation de Hamilton-Jacobi-Bellman. Introduisons la fonction V définie de la façon suivante :

$$V(x(\tau), \tau) = \min_{u(\tau), \dots, u(T-1)} \sum_{t=\tau}^{T-1} g(x(t), u(t), t) + h(x(T))$$

En sortant le premier terme de la somme, on constate que l'on peut écrire la formule de récurrence suivante :

$$V(x(\tau), \tau) = \min_{u(\tau)} g(x(\tau), u(\tau), \tau) + V(x(\tau+1), \tau+1)$$

Par conséquent pour résoudre ce problème, on détermine V de manière récursive en partant de la condition finale $V(x(T), T) = h(x(T))$. Cette méthode est efficace lorsque le nombre de variables d'états est petit (inférieur à 3 ou 4) sinon le temps de calcul et l'encombrement de la mémoire devient trop important.

3.2.5 Recuit simulé

Le recuit simulé est une méthode d'optimisation combinatoire issue de la physique qui s'inspire des techniques de sidérurgie pour faire refroidir le métal. Lors du refroidissement d'un métal, les atomes s'organisent en structures cristallines. Si la vitesse de refroidissement est rapide, le métal sera fragile. Pour éviter cette fragilisation, on effectue un refroidissement par palier. On descend donc la température T , puis on attend un certain temps avant de refroidir à nouveau. L'algorithme du recuit simulé proposé par Metropolis en 1953 (Metropolis et al. (1953)) pour minimiser une fonction s'appuie sur le critère de Monte Carlo. Ce critère repose sur une probabilité d'acceptation. On génère une solution j autre que la solution initiale i , la fonction objectif f s'en trouve modifier. On calcule la différence entre la valeur des deux fonctions objectifs $\delta f_{ij} = f(j) - f(i)$. Si cette différence est négative, alors la solution j est conservée sinon on la conserve avec une probabilité $\exp(-\frac{\Delta f_{ij}}{T})$. L'algorithme classique du recuit simulé se présente sous cette forme (Tableau 3.1) :

Les principaux inconvénients du recuit simulé résident dans le choix des nombreux paramètres, tels que la température initiale, la loi de décroissance de la température, les critères d'arrêt ou la longueur des paliers de température (Siarry & Dreyfus (1989), Reeves (1995), Saït & Youssef (1999)). Ces paramètres sont souvent choisis de manière empirique. Néanmoins,

TABLE 3.1 – Algorithme du recuit simulé

```
 $i = i_0$    où  $i_0$  solution initiale  
 $T = T_0$    où  $T_0$  température initiale  
Tant que critère d'arrêt n'est pas vérifié  
  Tant que fin de palier non atteint  
    générer une nouvelle solution  $j$   
     $\Delta f_{ij} = f(j) - f(i)$   
    Si  $\Delta f_{ij} < 0$  alors solution  $j$  acceptée  
    Sinon  
      Soit  $p$  probabilité tirée aléatoirement dans une loi uniforme  $U[0, 1]$   
      Si  $p < \exp(\frac{-\Delta f_{ij}}{T})$   
        alors solution  $j$  acceptée et  $T$  diminue  
      Fin si  
    Fin si  
  Fin tant que  
Fin tant que
```

c'est une méthode qui est facile à programmer à condition que la fonction objectif soit facilement et rapidement évaluable. Cette méthode est très souvent utilisée dans le domaine de l'industrie notamment sur l'étude de la conception de circuits électroniques et de réseaux informatiques.

3.3 Les algorithmes génétiques

Pourquoi choisir les algorithmes génétiques comme méthode d'optimisation, quelles sont les différences par rapport aux autres méthodes ?

Les algorithmes génétiques sont des algorithmes d'optimisation stochastiques fondés sur les mécanismes de la sélection naturelle et de la génétique (Darwin (1859)). Ils ont été initialement développés par John Holland (Holland (1962*a,b*)) et vulgarisés par Goldberg (Goldberg (1989)). Définissons tout d'abord quelques termes avant de décrire le fonctionnement des algorithmes génétiques :

- *Individu/chromosome/séquence* : une solution potentielle du problème. Un chromosome est une entité organisatrice servant de support à l'information génétique. Dans les algorithmes génétiques, une séquence correspondant à un individu. Le chromosome représente une chaîne (ensemble de gènes) qui code une solution. Les différentes versions d'un même gène sont appelées allèles. Dans notre exemple (tableau 3.2), les individus sont représentés par un chromosome constitué de 4 gènes. L'allèle 1 représente l'expression du caractère et l'allèle 0 représente son absence.

Individu	Chromosome			
A	1	0	0	0
B	0	1	1	0
C	0	1	0	0
D	0	0	1	0

TABLE 3.2 – Exemple de représentation d'un individu

- *Population* : un ensemble de chromosomes ou de points de l'espace de recherche,
- *Environnement* : l'espace de recherche,
- *Fonction de fitness/ d'évaluation* : la fonction que nous cherchons à maximiser.

Lerman et Ngouenet (Lerman & Ngouenet (1995)) ont mis en évidence 3 grands avantages de l'utilisation des algorithmes génétiques :

- Les algorithmes génétiques travaillent sur une population d'individus c'est à dire sur plusieurs solutions à la fois, au lieu d'un individu unique c'est à dire une seule solution.
- Les algorithmes génétiques n'utilisent que les valeurs de la fonction étudiée, et non pas sa dérivée,
- Les algorithmes génétiques utilisent des règles de transition probabilistes, et non déterministes, ce qui permet de s'affranchir des optima locaux.

Leurs champs d'application sont très vastes :

- optimisation de fonctions (De Jong (1980))
- finance (Pereira (2000))
- théorie du contrôle optimal (Krishnakumar & Goldberg (1992), Marco et al. (1996), Michalewicz et al. (1992))
- génétique (Carlborg et al. (2000), Nakamichi et al. (2001), Ljungberg et al. (2004))
- conception d'antennes lentilles (Godi et al. (2005))

Les algorithmes génétiques sont parfois fusionnés à d'autres algorithmes stochastiques comme par exemple avec le recuit simulé (Davis (1987), Mahfoud & Goldberg (1995)) avec l'algorithme du gradient (Oulladji et al. (2003)), avec les réseaux de neurones (Montana & Davis (1989), Forrest & Mitchell (1993)).

La description du fonctionnement général des algorithmes génétiques ainsi que des différents opérateurs et des améliorations possibles de ces algorithmes a été faite par la combinaison de différents sites internet :

- www.a525g.com/intelligence-artificielle/algorithmes-genetique.htm
- www.recherche.enac.fr/opti/papers/thesis/HABIT/main002.html
- www.recherche.enac.fr/opti/papers/thesis/thesend.pdf

Pour utiliser les algorithmes génétiques, nous devons disposer des 5 éléments suivants :

1. Un principe de codage des individus de la population. Cette étape associe à chaque individu une structure de données. Elle se place généralement après une phase de modélisation mathématique du problème traité. Le codage binaire a été très utilisé à l'origine. De nos jours, les codages réels sont largement utilisés.
2. Un générateur de la population initiale. Ce mécanisme doit être capable de produire une population d'individus non homogène qui servira de base pour les générations futures.
3. Une fonction à optimiser. Celle-ci retourne une valeur pour chaque individu appelée *fitness* ou *fonction d'évaluation de l'individu*.
4. Des opérateurs permettant de diversifier la population au cours des générations et d'explorer l'espace de recherche. L'*opérateur de croisement* recompose les gènes d'individus existant dans la population, l'*opérateur de mutation* a pour but de garantir l'exploration de l'espace des solutions.
5. Des paramètres de dimensionnement : taille de la population, nombre total de générations ou critère d'arrêt, probabilités d'application des opérateurs de croisement et de mutation.

Le fonctionnement général des algorithmes génétiques se déroule de la manière suivante (figure 3.1) : une population initiale d'individus est créée de manière aléatoire. Pour passer à la génération suivante, des couples de parents P_1 et P_2 sont sélectionnés en fonction de leur fitness. On leur applique l'opérateur de croisement avec la probabilité P_c .

A partir de ce croisement, des couples d'enfants C_1 et C_2 sont créés. En parallèle, d'autres individus P sont sélectionnés selon leur fitness. On leur applique l'opérateur de mutation avec la probabilité P_m et on obtient de nouveaux individus P' . La fitness des individus C_1 , C_2 , P' ainsi créés (par croisement ou par mutation) est évaluée avant qu'ils soient insérés dans la population de la génération suivante. Ce processus est itéré plusieurs fois. Le critère d'arrêt est variable : le nombre de générations est fixé a priori ou arrêt de l'algorithme quand la population n'évolue plus. Les opérateurs de sélection, de croisement et de mutation jouent un rôle prépondérant dans la réussite des algorithmes génétiques. Chacun de ces opérateurs agit selon divers critères qui lui sont propres (valeur sélective des individus, probabilité d'activation de l'opérateur). Nous allons décrire, dans la partie suivante, les particularités de chacun de ces différents opérateurs.

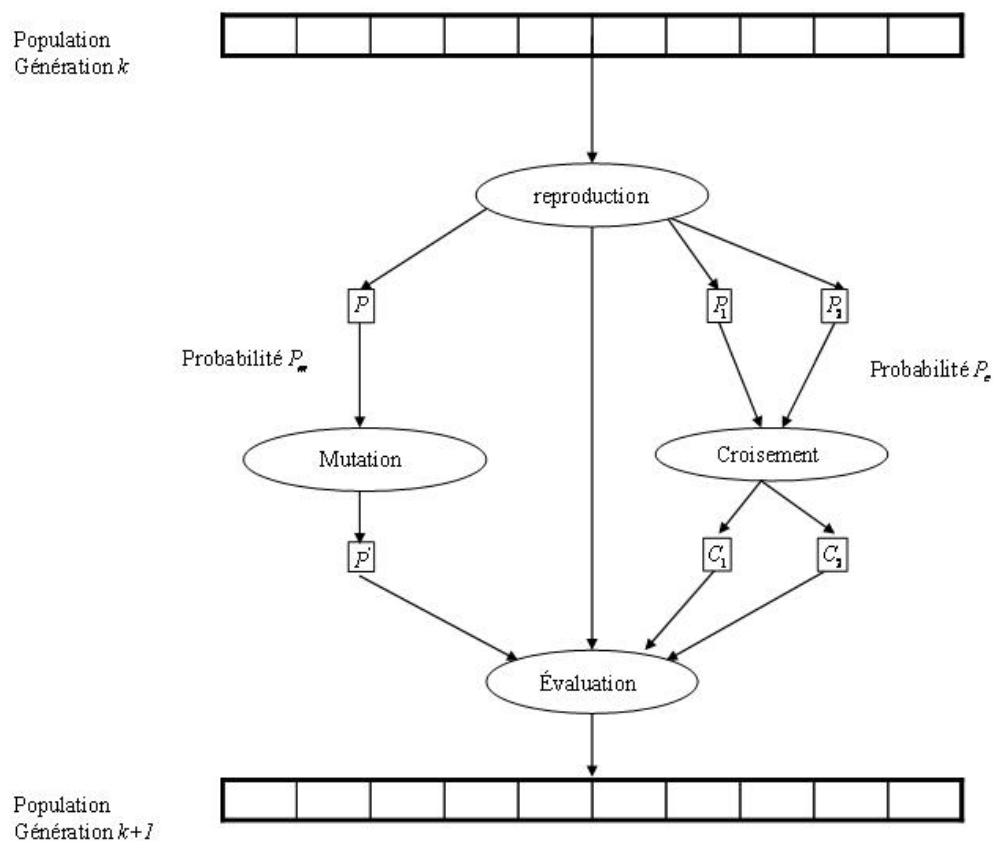


FIGURE 3.1 – Fonctionnement général des algorithmes génétiques
(source : www.recherche.enac.fr)

3.4 Les différents opérateurs des algorithmes génétiques

3.4.1 Opérateur de sélection

Cet opérateur est peut être le plus important des différents opérateurs présents dans les algorithmes génétiques. En effet, les individus se reproduisent, survivent ou meurent sous l'action des opérateurs de sélection qui interviennent à deux endroits : sélection pour la reproduction et sélection pour le remplacement. Il existe deux groupes de sélection : la « sélection proportionnelle » et la « sélection par rang ». Dans la sélection proportionnelle, on compte la *proportionate selection* (Holland (1975)), *stochastic remainder selection* (Booker (1982) et Brindle (1981)), *roulette wheel selection* (Goldberg (1989)), *stochastic universal sampling selection* (Baker (1987), Grefenstette & Baker (1989)). Dans la sélection par rang, on liste la sélection par tournoi (Brindle (1981), Goldberg (1989)), *truncation selection* Muhlenbein & Schlierkaamp-Voosen (1993), *linear ranking selection* Baker (1985). Nous allons décrire les méthodes les plus couramment utilisées.

Sélection par roulette

La méthode la plus connue et la plus utilisée est la roue de la loterie biaisée (*roulette wheel*) de Goldberg Goldberg (1989). Cette méthode consiste à associer à chaque individu i (chromosome) une probabilité P_i proportionnelle à sa fitness $f(i)$. Cette probabilité est calculée de la manière suivante : $P_i = f(i) / \sum f(i)$. Chaque individu i est sélectionné avec

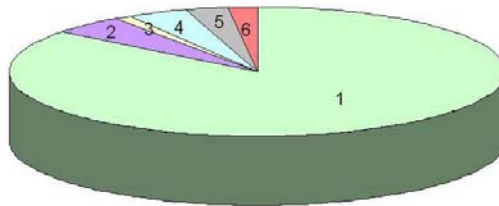


FIGURE 3.2 – Exemple de sélection par roulette de wheel (source :<http://sis.univ-tln.fr>)

la probabilité P_i , les « bons » individus auront ainsi plus de chance d'être sélectionnés, et ce plusieurs fois, d'autres le seront moins ou voire même jamais. Par exemple dans la figure 3.2, le chromosome 1 aura plus de chance d'être sélectionné et donc reproduit que le chromosome 3.

Sélection par tournoi

La méthode de sélection par tournoi est plus connue par son terme anglo-saxon *tournament selection*. Cette méthode augmente les chances des « mauvais » individus d'être sélectionnés pour conserver une certaine diversité dans la population. Cette méthode sélectionne k individus de manière aléatoire ou par la méthode « *roulette wheel* ». Ensuite un individu est choisi parmi les k individus sélectionnés, le meilleur selon la fitness. Ce processus s'effectue autant de fois qu'il y a d'individus à sélectionner. Le nombre k représente la taille du tournoi. Winter et al. (1995) ont démontré que les meilleurs résultats étaient obtenus avec une valeur de k égale à 2. Cette méthode permet à certains individus d'être sélectionnés plusieurs fois entraînant la transmission de leurs gènes à la nouvelle génération.

TABLE 3.3 – Exemple de sélection par rang (source :<http://sis.univ-tln.fr>)

Chromosomes	1	2	3	4	5	6	7	8
Probabilités Roulette wheel	79%	4%	0.5%	3%	9%	1%	2%	1.5%
Rang	8	6	1	5	7	2	4	3
Probabilités Sélection par rang	22%	17%	3%	14%	19%	6%	11%	8%

Sélection par rang

La sélection précédente doit faire face à des problèmes lorsque la valeur d'adaptation des chromosomes varie de façon importante. Si la meilleure fonction d'évaluation d'un chromosome représente 90% de la roulette, alors les autres chromosomes auront très peu de chance d'être sélectionnés et l'on arrivera à une stagnation de l'évolution. La sélection par rang trie d'abord la population par fitness (valeur d'adaptation). Ensuite, chaque chromosome se voit associé un rang en fonction de sa position. Ainsi le plus mauvais chromosome aura le rang 1, le suivant 2, et ainsi de suite jusqu'au meilleur chromosome qui aura le rang N , pour une population de N chromosomes. La sélection par rang d'un chromosome est la même que par roulette, mais les proportions sont en relation avec le rang plutôt qu'avec la valeur de l'évaluation. Le tableau 3.3 fournit un exemple de sélection par rang. Avec cette méthode de sélection, tous les chromosomes ont une chance d'être sélectionnés. Cependant, la convergence vers la bonne solution est lente puisque les meilleurs chromosomes n'ont guère plus de chance que les plus mauvais d'être sélectionnés.

Elitisme

Les individus sont triés selon leur fitness. Une partie de la population est sélectionnée, en général la moitié supérieure est conservée. Les individus les plus forts participent donc à l'amélioration de la population. Cette méthode converge rapidement, cependant elle diminue la diversité génétique des individus et est donc très sensible à la présence d'optimum locaux. Il est pourtant nécessaire de conserver une certaine diversité génétique. En effet, la combinaison de deux parents moyens peut donner un individu très intéressant.

Sélection steady-state

Une grande partie de la population initiale doit survivre à la prochaine génération. Par conséquent, à chaque génération, quelques chromosomes sont sélectionnés (les meilleurs selon la fitness) pour créer des chromosomes fils. Ensuite, les chromosomes les plus mauvais sont retirés et remplacés par les nouveaux. Le reste de la population survit à la nouvelle génération.

3.4.2 Opérateur de reproduction

La reproduction est composée de deux phases : le croisement et la mutation. Le croisement est un opérateur sexué, deux parents sont nécessaires pour créer deux chromosomes fils. L'opérateur de mutation est asexué, il suffit d'un chromosome parent pour générer un chromosome enfant.

L'opérateur de croisement

Cet opérateur est appliqué sur la population sélectionnée. Les couples sont créés de manière aléatoire et se reproduisent pour générer la nouvelle génération. Les chromosomes (ensemble de paramètres) des parents sont alors copiés et recombinaison de façon à former deux descendants possédant des caractéristiques issues des deux parents. Il existe plusieurs croisements possibles : le croisement en 1 ou k points et le croisement uniforme.

Le croisement en 1 point

A partir de deux individus, on obtient deux nouveaux individus (enfants) qui héritent de certaines caractéristiques de leurs parents. La population sélectionnée fait office de parents pour créer la nouvelle génération. La probabilité de croisement représente la fréquence à laquelle les croisements sont appliqués. Dans les cas extrêmes, si la probabilité de croisement est de 0% alors la nouvelle génération est la copie de la population précédente. Si la probabilité est de 100%, aucun des individus de la population précédente n'est conservée dans la nouvelle population. Nous choisissons au hasard un point de coupure k tel que $k \geq 1$ et $k < n$ (le nombre d'allèles). Un premier descendant est construit en concaténant les k premiers allèles du premier parent au $n - k$ allèles du second. De même un deuxième descendant est construit en concaténant aux k allèles du second parent les $n - k$ allèles du premier parent (Figure 3.3). Le croisement à k points est une généralisation du croisement à un point.

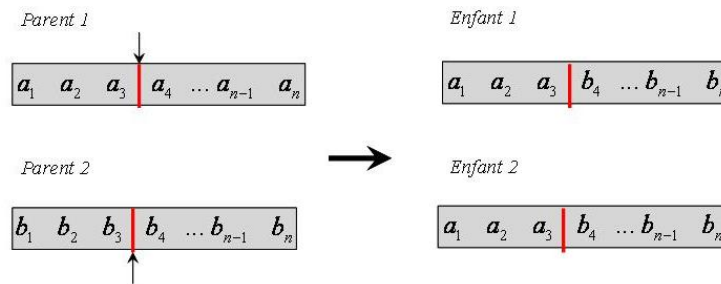


FIGURE 3.3 – Exemple du principe du croisement à un point

Le croisement uniforme

La mise en oeuvre de ce procédé consiste à définir de manière aléatoire un « masque », c'est-à-dire une chaîne de même longueur que les chromosomes des parents sur lesquels il sera appliqué. Ce masque permet de savoir pour chaque locus de quel parent les fils devront hériter du gène (figure 3.4). Si face à un locus le masque présente un 0, le fils héritera du gène du parent n°1, s'il présente un 1 il héritera du gène du parent n°2. La création du fils n°2 se fait de manière symétrique : si pour un gène donné le masque indique que le fils n°1 devra recevoir celui du parent n°1 alors le fils n°2 recevra le gène du parent n°2. A l'opposé, si le fils n°1 reçoit le gène du parent n°2 alors le fils 2 recevra le gène du parent n°1. L'opérateur de croisement favorise ainsi l'exploration de l'espace des solutions.

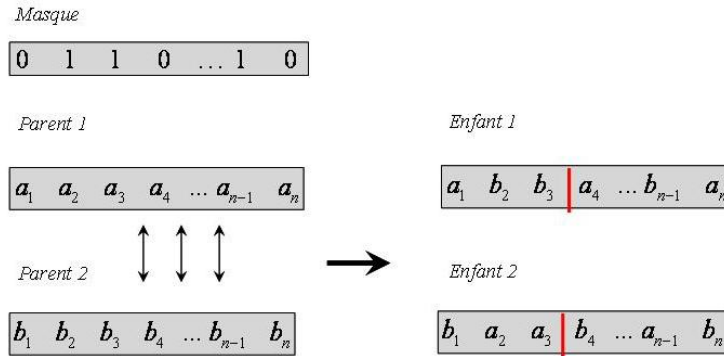


FIGURE 3.4 – Exemple du principe du croisement uniforme

L'opérateur de mutation

La mutation génère des « erreurs » de copie, afin de créer un nouvel individu qui n'existait pas auparavant. La valeur allélique d'un gène est changée avec une probabilité p_m très faible (comprise entre 0.01 et 0.001). Le but est de créer de nouveaux individus et d'éviter aux algorithmes génétiques de converger vers des extrema locaux de la fonction objectif. Si elle génère un individu faible, l'individu est éliminé. S'il n'y a pas de mutation, le fils est inséré dans la nouvelle population sans changement. Cet opérateur modifie de manière aléatoire les caractéristiques des individus, ce qui permet d'introduire et de maintenir la diversité au sein de la population. De plus, cet opérateur permet de ne pas avoir de dérive génétique. Sans lui, certains allèles favorisés par le hasard se répandent au détriment des autres. Le fait que l'opérateur de mutation puisse entraîner aléatoirement des changements au niveau de n'importe quel locus permet d'amoindrir ce phénomène.

D'autre part, cet opérateur permet de limiter les risques d'une convergence prématurée causée par exemple par une méthode de sélection élitiste imposant à la population une pression sélective trop forte. En effet, dans le cas d'une convergence prématurée, on se retrouve avec une population dont tous les individus sont identiques mais ne sont que des optimums locaux. Tous les individus étant identiques, le croisement ne changera rien à la situation. L'évolution est bloquée et par conséquent, l'optimum global ne sera jamais atteint. La mutation permet d'effectuer des inversions de manière aléatoire et d'éviter ainsi cette situation de convergence prématurée.

3.4.3 Les paramètres de dimensionnement

Les opérateurs de l'algorithme génétique sont guidés par un certain nombre de paramètres fixés à l'avance : la taille de population, le nombre de générations à simuler, la probabilité d'application des opérateurs. La valeur de ces paramètres influence la réussite ou non des algorithmes génétiques.

La taille de la population

La taille de la population est proportionnelle à la diversité (Winter et al. (1995)). En effet, si la taille de population est réduite, l'exploration de l'espace de recherche est insuffisante. La solution obtenue ne sera qu'un optimum local dû au manque de diversité. En revanche,

si la taille de la population est trop grand, le temps de calcul de l'algorithme peut s'avérer prohibitif.

La probabilité de croisement p_c

Cette probabilité dépend de la forme de la fonction de fitness. Son choix est en général heuristique. Plus elle est élevée, plus la population subit de changements importants. Les valeurs généralement admises sont comprises entre 0,5 et 0,9.

La probabilité de mutation p_m

Cette probabilité est généralement faible puisqu'une probabilité élevée risque de conduire à une solution sous optimale. Elle est comprise généralement entre 0,001 et 0,01.

3.5 Les améliorations classiques

Les processus de sélection sont très sensibles aux écarts de fitness. Par exemple, un très bon individu risque d'être reproduit trop souvent et pourrait induire l'élimination de certains individus, ce qui amènerait à une population homogène contenant un seul type d'individu. Par exemple dans la figure 3.5, l'individu correspondant au sommet S_2 risque d'être le seul représentant pour la génération suivante. Les autres individus ont une fitness faible. Seule la mutation pourrait aider à atteindre l'objectif global (individu correspondant au sommet S_1 après de nombreuses tentatives. Afin d'éviter ce cas de figure, il existe des principes (*sca-*

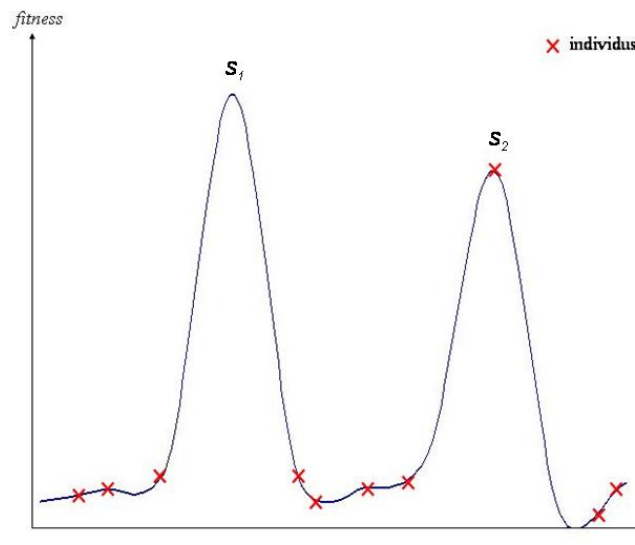


FIGURE 3.5 – Exemple de risque de reproduction d'un seul individu

ling, sharing) qui empêchent les individus « forts » d'éliminer complètement les individus « faibles » pour éviter ce phénomène de « niche ».

3.5.1 Scaling

Le scaling ou mise à l'échelle transforme la fitness de chaque individu afin de réduire ou d'amplifier les différences entre les individus. La sélection ne se fait plus sur la fitness réelle mais sur la fitness transformée. Pour modifier la fitness d'origine, différentes transformations peuvent être envisagées linéaires ou exponentielles.

3.5.2 Sharing

Le sharing est une technique de partage inspirée d'un phénomène naturel de création de niche écologique. Pour éviter le rassemblement des individus autour d'un sommet dominant, le sharing pénalise les fitness en fonction du taux d'agrégation de la population dans le voisinage d'un individu (figure 3.6). Pour ce faire, on introduit une notion de distance de dissimilarité

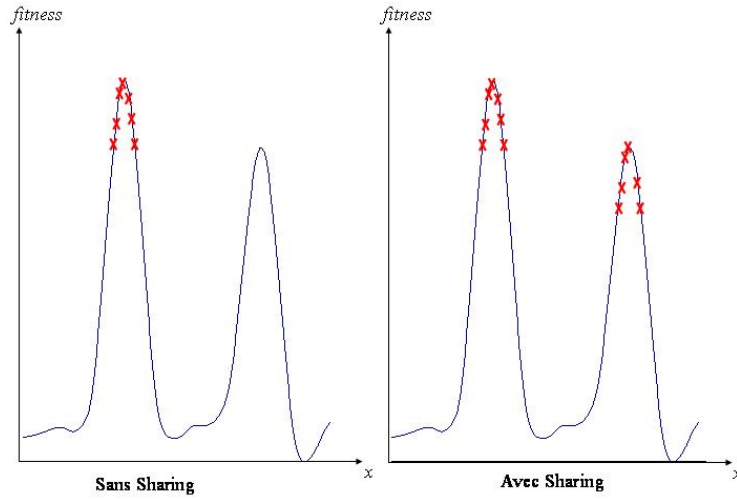


FIGURE 3.6 – Principe du Sharing

d entre les individus. La fitness de chaque individu est calculée en fonction de son voisinage.

$$f'(i) = \frac{f(i)}{c_i}$$

avec

$$c(i) = \sum_{j=1}^N Sh(d(i, j))$$

$$Sh(d) = \begin{cases} 1 - \frac{d}{\sigma_{share}} & \text{si } d \leq \sigma_{share} \\ 0 & \text{sinon} \end{cases}$$

où σ_{share} correspond à la largeur du voisinage, c'est-à-dire qu'il permet de délimiter le voisinage d'un point. Après cette modification de la fitness, la sélection est effectuée.

Goldberg & Richardson (1987) ont montré que l'introduction de cette technique améliore la performance des algorithmes génétiques.

3.5.3 Parallélisation

Les algorithmes génétiques sont des algorithmes coûteux en temps calcul d'où l'intérêt du calcul parallèle. La parallélisation des algorithmes génétiques peut se faire assez facilement grâce à l'indépendance entre les différentes opérations génétiques. Il existe pour cela différentes méthodes. L'étape la plus coûteuse en temps de calcul est l'évaluation de la performance (fitness) de toute la population. Les calculs de fitness sont indépendants d'un individu à l'autre ce qui permet le calcul parallèle. Le gain de temps obtenu dépend alors du problème considéré.

Le modèle en îlots

Cette méthode consiste à diviser la population en sous-populations (figure 3.7). Chaque sous-population évolue suivant son algorithme génétique auquel vient s'ajouter une étape de migration. Chaque sous-population envoie ses meilleurs individus, soit vers les populations voisines soit dans un « pool » commun. Chaque sous-population reçoit ensuite des individus soit envoyés par ses voisins soit pêchés dans le « pool » commun.

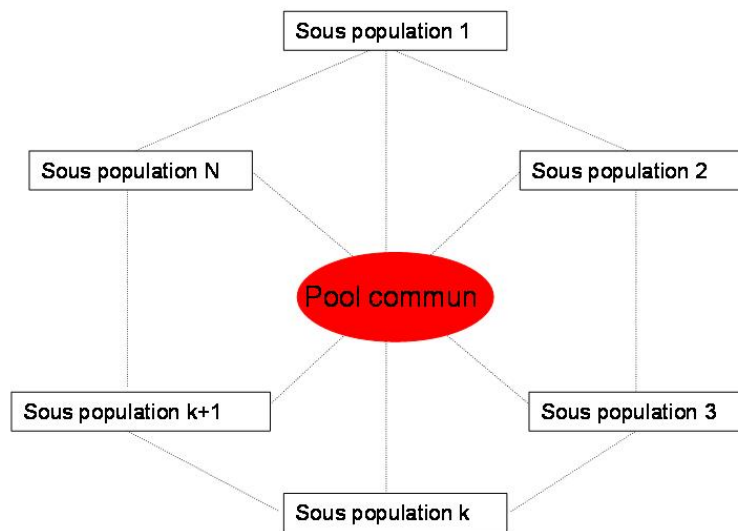


FIGURE 3.7 – Fonctionnement du modèle en îlots

Ce phénomène de migration permet de préserver la diversité génétique le plus longtemps possible. Cependant, si un grand nombre d'individus de chaque sous-population migre de manière fréquente, la diversité globale sera alors très vite diminuée. Dans le cas contraire, si la migration est peu fréquente, nous nous trouvons face à une convergence prématurée des sous-populations (Bubak et al. (1997)). Ce phénomène rajoute deux paramètres (taux de migration et la fréquence de cette dernière) difficiles à régler.

Modèle à population distribuée

Cette méthode est basée sur une approche « multi-agent ». La population initiale est générée par un agent « Interface » (figure 3.8). Cet agent partitionne cette population en sous populations de taille identique et les distribue à des agents « Espèces ». De là chaque

population évolue en fonction de son AG. Ce modèle génère lui aussi une nouvelle dynamique de l'algorithme, différente de la précédente, avec l'apparition de « niches écologiques » distribuées par « zones géographiques », et permettant donc également la découverte d'optima multiples. L'agent « Interface » reçoit et traite toutes les informations des agents « Espèces ». Cet agent stoppe cette méthode quand il considère que l'un des agents « Espèces » lui a fourni la solution optimale.

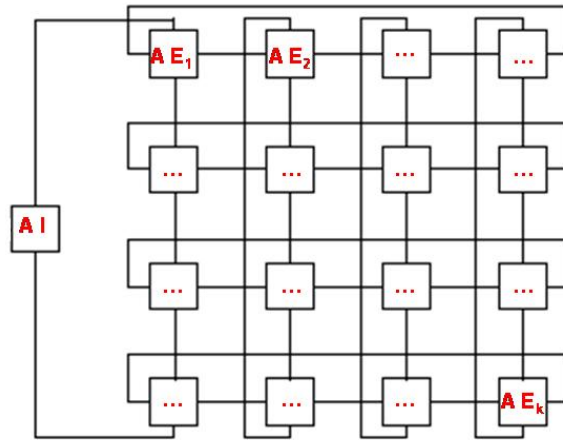


FIGURE 3.8 – Fonctionnement du modèle à population distribuée

L'algorithme génétique le plus connu qui utilise cette méthode est *Distributed Guided Genetic Algorithm* (DGGA) initialement proposé par Ghédira & Jlifi (2002). DGGA a été modifié principalement au niveau des opérateurs de croisement et de mutation par Bouamama & Ghédira (2005).

Les approches parallèles

Grefenstette (1981) a étudié de nombreuses implémentations de parallélisation des algorithmes génétiques que nous développons par la suite :

- Maître-Esclave synchrone,
- Maître-Esclave semi-synchrone,
- Approche distribuée concurrente asynchrone.

1. Maître-Esclave synchrone

Il s'agit de paralléliser l'étape d'évaluation. Un processus Maître gère l'algorithme lui-même (sélection/remplacement et opérateurs génétiques), et envoie les calculs de performance à des processus Esclaves (figure 3.9). Cette méthode de parallélisation comporte deux désavantages : synchronisme du fonctionnement et dépendance du processus Maître.

2. Maître-Esclave semi-synchrone

Cette méthode résout le premier inconvénient de la méthode synchrone en insérant et en sélectionnant des individus à la volée alors que les processus Esclaves effectuent

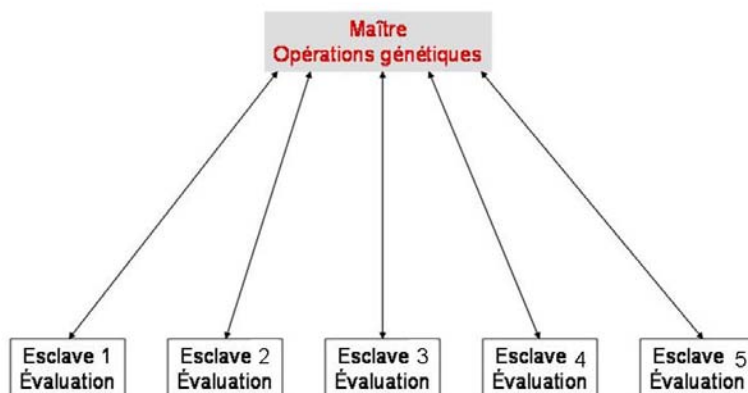


FIGURE 3.9 – Schéma du Maître-Esclave synchrone

leur travail (figure 3.10). L'inconvénient de la dépendance du processus Maître reste cependant présent (Laurens (1995)).

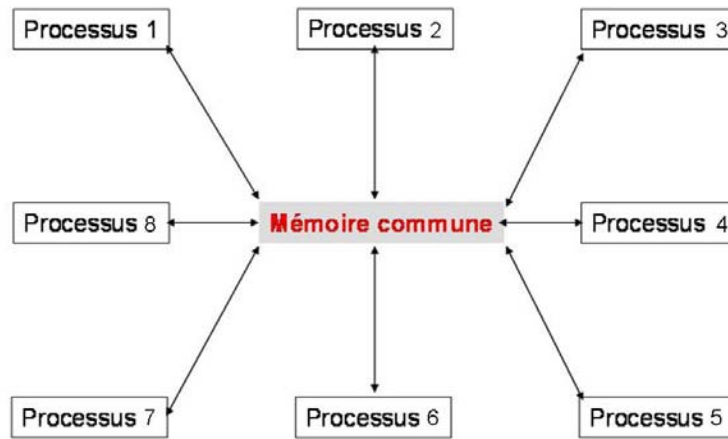


FIGURE 3.10 – Schéma du Maître-Esclave asynchrone

3. Approche distribuée concurrente asynchrone

Une manière de résoudre ce problème est d'utiliser une méthode asynchrone. Pour cela, on dispose de k processus qui exécutent les opérations génétiques et l'évaluation indépendamment les uns des autres en utilisant une mémoire partagée Talbi & Bessière (1991a,b). L'implémentation de cette approche est plus difficile à mettre en place que les méthodes Maître Esclave.

3.6 Convergence

Holland (1975) a prouvé que même dans des espaces larges et complexes, les algorithmes génétiques convergent sous certaines conditions vers des solutions globalement optimales autrement dit la population de cet algorithme se concentre autour d'un optimum global. Puis en 1991, Davis & Principe (1991) ont démontré la convergence de l'algorithme génétique simple. Les trois opérateurs des algorithmes génétiques (sélection, croisement et mutation) jouent un rôle prépondérant dans la convergence vers l'optimum global. En effet, la sélection agit comme un opérateur de « rassemblement » alors que le croisement agit comme un opérateur d'exploration. Ces deux opérateurs favorisent la concentration de la population dans des zones de l'espace de recherche où la fonction d'évaluation est assez élevée. Si cette concentration est faite prématurément un risque de convergence vers un optimum local est élevé. On parle alors de perte de diversité génétique c'est pourquoi l'opérateur mutation intervient pour éviter cette perte de diversité. CERF (1994), Davis & Principe (1991) et Nix & Vose (1992) ont montré les effets et l'efficacité de ces opérateurs. Le réglage des paramètres des algorithmes génétiques est assez délicat mais primordiale pour une convergence vers un optimum global dans un temps « correct ». Par exemple, si la taille de la population est trop petite, l'algorithme va évoluer très probablement vers un optimum local. Dans le cas contraire, si la population est trop importante le temps de convergence sera excessif. La taille de la population dépend des méthodes utilisées (sélection, opérateurs génétiques), du nombre de variables,

de la fonction objective et de la puissance de calcul disponible. A chaque problème correspond son algorithme génétique.

3.7 Conclusion

Les algorithmes génétiques ont pour but d'obtenir une solution à un problème d'optimisation lorsqu'il n'existe pas ou qu'on ne connaît pas de méthode exacte pour le résoudre. Les algorithmes génétiques sont plus performants que les algorithmes classiques lorsque l'espace de recherche est important. Néanmoins, certains paramètres comme la taille de la population ou le taux de mutation sont parfois difficiles à déterminer et plusieurs ajustements sont nécessaires. De plus, le choix de la fonction objective est très important puisqu'elle doit prendre en compte les « vrais » paramètres du problème. Enfin, la solution obtenue même après de nombreuses générations n'est pas forcément la solution optimale mais une solution acceptable qui s'en approche le plus. De nombreuses méthodes hybrides ont été adaptées aux algorithmes génétiques. Néanmoins ces méthodes augmentent le nombre de paramètres initiaux à ajuster. Le choix d'une méthode plus qu'une autre est vraiment heuristique et dépend réellement des besoins de l'utilisateur (précision de la solution, temps d'exécution). Dans ce travail, le choix s'est porté sur l'algorithme génétique couplé avec la technique du sharing.

Deuxième partie

Modèle déterministe

Chapitre 4

Modèle déterministe

4.1 Résumé

Un modèle déterministe a été développé afin d’optimiser simultanément la sélection sur deux caractères indépendants, un caractère monogénique pour la résistance aux maladies et un caractère de production polygénique. Pour prendre en compte le processus de sélection, une modélisation dynamique a été utilisée. Ce modèle décrit l’évolution de la population au cours du temps sous l’influence de phénomènes naturels (vieillessement, mort) et artificiels (sélection). Cette description se fait à l’aide de paramètres (fixés par la nature), de variables de décision (taux de sélection, probabilités d’accouplements entre individus élites selon leur génotype, variables sur lesquelles on peut agir) et des variables d’état. Ces dernières sont des statistiques (fréquence, espérance, variance) décrivant les différentes classes de la population ayant les mêmes caractéristiques. Ces classes sont caractérisées par le sexe (s), l’âge (a), le génotype (g) au gène à sélectionner et la catégorie (c) (élite *vs.* non élite dans le cas des mâles et des femelles en âge de reproduire). L’état d’une classe est décrit par sa fréquence génotypique f_{sactg} et l’espérance de la distribution du caractère polygénique μ_{sactg} pour chaque génération t . L’objectif de la sélection est décrit à l’aide d’une fonction objective qui dépend des paramètres et des variables du modèle. Par exemple, cette fonction peut être la fréquence du génotype favorable des femelles au sein de la population lors de la dernière année du processus. L’optimisation des variables de décision pour maximiser cette fonction objective se fait en utilisant un algorithme génétique. Compte tenu du caractère dynamique de l’optimisation, les variables de décision au cours du processus doivent être optimisées simultanément.

L’exemple de la résistance à la tremblante contrôlée par le gène PrP a été traité. Dans cette application, nous avons considéré que le gène possède deux classes d’allèles (R résistant, S sensible). Notre objectif était d’obtenir une population résistante tout en limitant la perte du progrès génétique. Cet objectif a été transcrit par la maximisation des fréquences des femelles porteuses du génotype de résistance à la fin du processus avec une pénalité sur la perte du progrès génétique sur le caractère polygénique. Le schéma de sélection considéré ici, prend en compte, pour les femelles, une sélection sur leur performance individuelle. Les mâles sont sélectionnés en deux étapes séquentielles : une sélection avant testage et une sélection après testage sur leur performance sur le caractère polygénique. Des accouplements raisonnés entre mâles et femelles élites sont considérés pour le renouvellement des mâles et des accouplements aléatoires sont pris en compte pour le renouvellement des femelles. Différents scénarios ont été traités en termes de fréquences génotypiques initiales et de contraintes sur la perte de

progrès génétique. Des simplifications de la stratégie optimale ont été envisagées pour obtenir une stratégie plus aisée à mettre en oeuvre.

Avec différentes fréquences initiales et sans contrainte sur la perte du progrès génétique sur le caractère polygénique, on constate que la stratégie optimale est la même : conservation totale des animaux porteurs du génotype de résistance, sélection drastique sur les animaux porteurs des autres génotypes. Cette stratégie permet de fixer le génotype de résistance très rapidement au détriment du progrès génétique sur le caractère polygénique. Plus la fréquence du génotype de résistance est faible plus la perte sur le progrès génétique est importante. En augmentant cette fréquence la perte sur le progrès génétique diminue. La stratégie optimale se focalise sur l'augmentation très rapide du nombre d'individus porteur du génotype de résistance. En prenant une fréquence initiale faible du génotype de résistance (cas où la perte sur le progrès génétique est la plus importante) et en faisant varier la contrainte sur la perte du progrès génétique, on constate une modification de la stratégie optimale. En effet, avec une contrainte sur la perte du progrès génétique assez forte, la proportion d'animaux porteurs du génotype de résistance sélectionnés varie d'une génération à l'autre. Lors de la première génération, une forte proportion de ces animaux est conservée. Elle est diminuée à la génération suivante afin de compenser la perte du progrès génétique. Ce cheminement se fait sur toute la durée du schéma de sélection jusqu'à la fixation du génotype de résistance dans la population. Néanmoins la stratégie concernant les animaux porteurs du génotype défavorable est toujours aussi drastique. Les animaux de génotype "intermédiaire" sont conservés plus longtemps entraînant une atténuation de la perte du progrès génétique. Ce contraste entre ces stratégies montre bien que les objectifs initiaux - savoir où sont les priorités, quels sont les compromis que l'on accepte de faire - doivent être bien définis. Il semble cependant difficile d'appliquer de telles stratégies optimales (taux de sélection variant à chaque génération et pour chaque classe). Après une analyse des résultats obtenus, l'éventualité de certaines simplifications apparaissent comme pouvant amener à des stratégies proches de l'optimum. Certains taux de sélection peuvent être considérés constants au cours du processus. Un résultat convenable voir optimal peut être atteint, tout dépend des conditions initiales et des priorités définies.

Le modèle déterministe et dynamique, développé ici, permet de s'adapter à de nombreuses situations. En effet, il considère des générations chevauchantes, prend en compte différents types de sélection (phénotypique, génotypique ou combinée génotype-phénotype) quel que soit le sexe (mâle et femelle) et donne la possibilité de considérer soit des accouplements raisonnés soit des accouplements aléatoires. Cette modélisation permet de concevoir des stratégies complexes mais également de voir si certaines simplifications sont envisageables afin de pouvoir les appliquer aisément en pratique. Ce modèle permet de considérer plusieurs allèles. Néanmoins ce nombre doit être limité. En effet, en prenant en compte un grand nombre d'allèles, le nombre de paramètres croît de manière exponentielle et l'optimisation devient difficile.

Cette méthodologie peut être extrapolée à la modélisation de la sélection de gènes affectant des caractères d'intérêt zootechnique comme par exemple la prolificité des brebis ou la conformation des carcasses d'agneaux. Cette approche originale peut être étendue à plusieurs gènes/QTLs en intégrant également l'effet d'épistasie. De plus, une autre évolution de ce modèle serait à envisager : la prise en compte de la réduction de la variance génétique additive. En effet, la variance additive diminue avec la sélection, ce phénomène n'a pas été pris en compte puisque nous avons supposé que cette dernière était constante au cours des générations. Ce modèle déterministe a évolué vers un modèle stochastique afin de considérer un nombre limité de reproducteurs (chapitre 5).

4.2 Article

Cet article a été publié dans la revue « Journal of Animal Science ».

JOURNAL OF ANIMAL SCIENCE
The Premier Journal and Leading Source of New Knowledge and Perspective in Animal Science

A dynamic deterministic model to optimize a multiple-trait selection scheme

A. D. Costard, Z. G. Vitezica, C. R. Moreno and J.-M. Elsen

J Anim Sci 2009.87:885-894.

doi: 10.2527/jas.2008-0898 originally published online Nov 21, 2008;

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://jas.fass.org/cgi/content/full/87/3/885>



American Society of Animal Science

www.asas.org

A dynamic deterministic model to optimize a multiple-trait selection scheme¹

A.D. Costard*, Z.G. Vitezica † ††, C. R. Moreno*, J.-M. Elsen*²

*INRA, SAGA station d'amélioration génétique animale, 31326 Castanet-Tolosan, France; † INSERM, U563, Centre de Physiopathologie de Toulouse Purpan, Toulouse, F-31300 France; †† Université Toulouse III Paul-Sabatier, UMRs563, Toulouse, F-31400 France.

Running head: optimizing multiple-trait selection

¹This work was supported by the EU project QLRT-2000-01733 'Scrapiefreesheep' and by the EU Network of Excellence 'Neuroprion'.

²Corresponding author: Anne.Devalle@toulouse.inra.fr

ABSTRACT

A mathematical approach was developed to model and optimize simultaneously selection on 2 traits, a quantitative trait with underlying polygenic variation and a monogenic trait (e.g. resistance to a disease). A deterministic model allows a global optimization of the selection scheme in order to maximize the frequency of the desired genotype for the monogenic trait while minimizing the loss of genetic progress on the polygenic trait. An additive QTL or gene was considered. Breeding programs with overlapping generations, different selection strategies for males and females, and assortative mating were modeled. A Genetic Algorithm was used to solve this optimization problem. This modeling approach may easily be adapted to a variety of underlying genetic models and selection schemes. This model was applied to an example where selection on the Prp gene for Scrapie resistance was introduced as an additional selection criterion in an already existing dairy sheep selection scheme.

Key words: genetic algorithm, global optimization, marker assisted selection, selection scheme

INTRODUCTION

Gene or QTL information could improve selection efficiency and allow the implementation of marker-assisted selection (MAS). MAS has been proposed to improve classical selection schemes for complex traits with low heritability or which are difficult or expensive to measure (Lande and Thompson, 1990; Meuwissen and van Arendonk, 1992; Dekkers and van Arendonk, 1998). Strategies to use QTL or gene information in MAS are mainly based on indices combining the EBV for QTL and the EBV for the residual polygenic effect. MAS increases the QTL frequency in the short term but can result in lower response in the long term (Lande and Thompson, 1990). Several studies (Gibson, 1994; Larzul *et al.*, 1997b; Pong-Wong and Woolliams, 1998) have explored the interactions between selection methods and time horizon. Dekkers and van Arendonk (1998) optimized selection on an identified QTL in a simple selection scheme (phenotypic selection, random mating) using an optimal control approach (Lewis, 1986). The method was extended by Chakraborty *et al.* (2002). The benefit of optimal selection on a single identified QTL and the differences between optimal and standard QTL selection were evaluated by Dekkers and Chakraborty (2001). Manfredi *et al.* (1998) optimized selection and mating on a QTL for a sex limited trait using sequential quadratic programming. In these studies, selection was for a single quantitative trait affected by 1 identified QTL of meaningful size as well as polygenes. The objective of this study is to optimize simultaneously selection on 2 traits, a monogenic trait controlled by an identified QTL or gene, and a 2nd independent polygenic trait that is unaffected by the major locus with a genetic algorithm (Holland, 1973; Goldberg, 1989; Mongeau, 2003). A deterministic and dynamic model is

presented for breeding programs including overlapping generations, different selection strategies for males and females, and non-random mating.

MATERIALS AND METHODS

Genetic Model

A population of infinite size with overlapping generations and different classes of animals defined by sex s ($s=1$ for males and $s=2$ for females), age a , and EBV categories c (elite, not elite) was considered. Animals from identical sex and age may be classified in different subsets according to their EBV mean and their use as reproducers (*e.g.*, the number of progeny per year). The QTL or gene has 2 alleles (B and b). Genotypes BB, Bb/bB, and bb are denoted $g1$, $g2$ and $g3$ respectively. The polygenic effects are assumed to follow the infinitesimal genetic model with a variance σ^2 . The identified QTL or gene is unlinked with genes that influence the quantitative trait. For generation t , each class of animals has a QTL or gene effect with a genotype frequency of f_{sactg} , and a polygenic effect denoted by the within class average EBV μ_{sactg} . Selection can involve the EBV of the quantitative trait, the identified QTL or gene or both of them. Random mating and assortative mating according to genotype or to EBV categories can be used. The deterministic and dynamic model includes different equations that correspond to a general scheme presented in Figure 1. The phenomena underlying the evolution of the population are ageing, selection and matings. The equations describing the scheme are as follow:

Without selection, after year $t=1$, the genotype frequency and the average EBV for an animal of sex s , category c and genotype g do not change between t and $t-1$

$$f_{sactg} = f_{s(a-1)c(t-1)g}$$

$$\mu_{sactg} = \mu_{s(a-1)c(t-1)g}$$

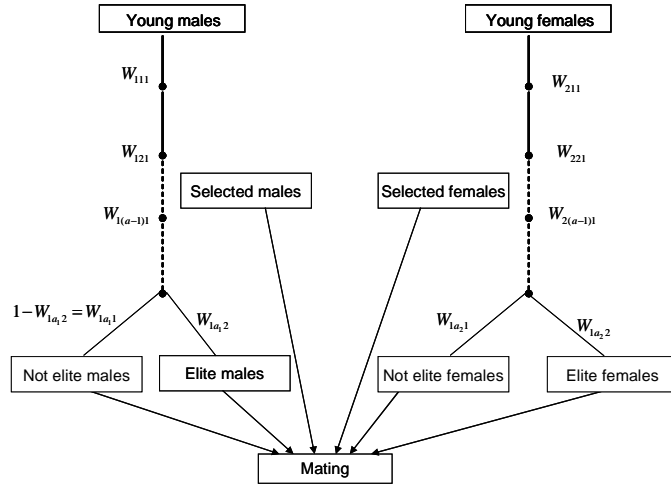


Figure 1. Selection scheme description with individual classes (rectangles), selection steps (•) and selection proportion W_{sac} (sex $s=1, 2$; age $a=1, \dots, a_s$; category $c=1, 2$)

Selection. Different selection steps after year $t=1$ can be considered for males and females (Figure 1). For the last selection step, males and females are distributed in 2 categories, elite and not elite. The selection rate W_{sac} can vary at each step for males and females. The expression for W_{sac} is:

$$W_{sac} = \sum_g w_{sactg} f_{s(a-1)c(t-1)g}$$

where w_{sactg} is the proportion of individuals with sex s , age a and category c at generation t with genotype g that are selected to become breeding animals. At each **step**,

candidates for selection can be evaluated through their phenotype, genotype or both of them.

Assuming that EBV's are normally distributed, w_{sactg} is equal to:

$$w_{sactg} = \phi(\beta_{2g} - \mu_{s(a-1)c(t-1)g}) - \phi(\beta_{1g} - \mu_{s(a-1)c(t-1)g})$$

where ϕ is the normal cumulative distribution function for a variance σ^2 , and β_{1g} and β_{2g} are lower and upper thresholds for genotype g (β_{1g} may be $-\infty$ and β_{2g} may be $+\infty$) defining the interval selected.

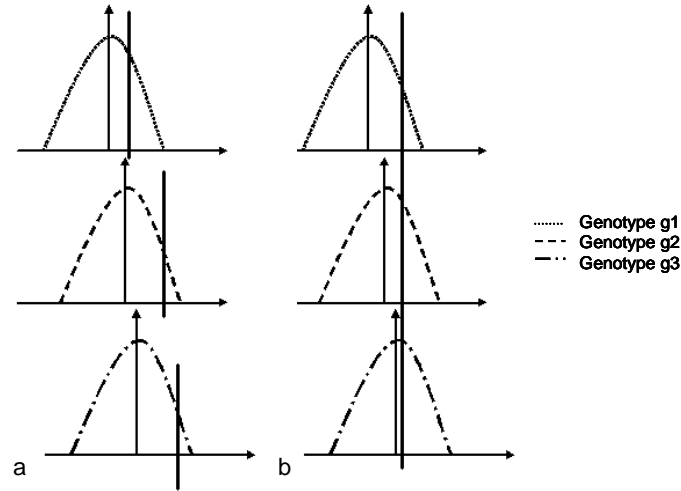


Figure 2. Selection threshold for genotype g1 (...), g2 (---) and g3 (-.-) under combined genotypic and phenotypic selection (a) and phenotypic selection (b)

When individuals are selected according to EBV and genotype information, the selection thresholds may vary according to the genotype (Figure 2a), and the state variables f_{sactg} and μ_{sactg} are given by

$$f_{sactg} = \frac{w_{sactg}}{W_{sac}} f_{s(a-1)c(t-1)g} \quad [1]$$

$$\mu_{sactg} = \mu_{s(a-1)c(t-1)g} + \frac{\varphi(\beta_{1g} - \mu_{s(a-1)c(t-1)g}) - \varphi(\beta_{2g} - \mu_{s(a-1)c(t-1)g})}{w_{sactg}} \rho \sigma \quad [2]$$

where $\varphi(x)$ is the normal density function, ρ is the accuracy or the correlation between true and the polygenic EBV and σ the genetic standard deviation. The phenotypic selection is a special case of these equations where β_1 and β_2 are the same for all genotypes, i.e., $\beta_{11} = \beta_{12} = \beta_{13} = \beta_1$ and $\beta_{21} = \beta_{22} = \beta_{23} = \beta_2$ (Figure 2b).

In the case of selection for a monogenic trait, the genotypes are ranked, because some genotypes are better than others from a biological point of view (*e.g.* favorable or unfavorable genotypes for a disease). Let $g[k]$ be the k^{th} ranked genotype. These constraints must be verified:

$$\sum_{k=1}^{r-1} f_{s(a-1)c(t-1)g[k]} < W_{sac} \text{ and } \sum_{k=1}^r f_{s(a-1)c(t-1)g[k]} \geq W_{sac} .$$

because there cannot be more individuals in those classes than in the whole population.

The selection proportions are:

$$w_{sactg[k]} = 1 \quad \text{if } k \leq r-1$$

$$w_{sactg[k]} = 0 \quad \text{if } k > r$$

$$w_{sactg[r]} = \frac{W_{sac} - \sum_{k=1}^{r-1} f_{s(a-1)c(t-1)g[k]}}{f_{s(a-1)c(t-1)g[r]}} \text{ otherwise}$$

The selection within the r^{th} genotype was at random. The genotype frequency f_{sactg} and the EBV mean μ_{sactg} of the population in generation t are obtained from equations [1] and [2].

Mating. Mating between different classes (selected, elite and not elite) of females and males can be considered (Figure 1). The probability for a progeny to be of genotype g , given that its parents carry the genotypes h (sire) and k (dam), is τ_{ghk} . The probability that a parent of sex s has genotype g is $F_{s(t-1)g}$ and is defined as:

$$F_{s(t-1)g} = \sum_{a,c} \delta_{sac} f_{sac(t-1)g}$$

where δ_{sac} is the proportion of individuals of age a and class c among the parents with sex s . This quantity depends on demographic constraints (*e.g.* reproductive age) and breeder decisions (*e.g.* relative use of progeny tested males). From the parameters τ_{ghk} and $F_{s(t-1)g}$, different matings (random or assortative) can be considered.

For random mating, none of the parents' characteristics (*e.g.* EBV or genotype) are considered in matings. The genotype frequency and the EBV mean for the newborn animals of age 1 and category 1 are:

$$f_{s11tg} = \sum_{h,k} \tau_{ghk} F_{1h(t-1)} F_{2k(t-1)}$$

$$\mu_{s11tg} = \frac{1}{2f_{s11tg}} \sum_{h,k} \tau_{ghk} \sum_{a_1,c_1} \sum_{a_2,c_2} \frac{\delta_{1a_1c_1} f_{1a_1c_1(t-1)h}}{F_{1h(t-1)}} \frac{\delta_{2a_2c_2} f_{2a_2c_2(t-1)k}}{F_{2k(t-1)}} (\mu_{1a_1c_1(t-1)h} + \mu_{2a_2c_2(t-1)k})$$

where $s=1$ for sire, $s=2$ for dam, a_1 and a_2 , and c_1 and c_2 are the age and category for the sire and dam.

For genotypic assortative mating, the mating may be organized according to the parents' genotypes. The optimization of these matings may be achieved by choosing

a value for α_{hkt} which is the proportion of newborn animals in generation t from a sire with genotype h and a dam with genotype k . This proportion is constrained by

$$\sum_k \alpha_{hkt} = F_{1h(t-1)}$$

$$\sum_h \alpha_{hkt} = F_{2k(t-1)}$$

In this situation, the genotype frequency and the EBV mean are given by

$$f_{s11tg} = \sum_{h,k} \tau_{ghk} \alpha_{hkt}$$

$$\mu_{s11tg} = \frac{1}{2f_{s11tg}} \sum_{h,k} \tau_{ghk} \alpha_{hkt} \sum_{a_1, c_1} \sum_{a_2, c_2} \frac{\delta_{1a_1c_1} f_{1a_1c_1(t-1)h}}{F_{1h(t-1)}} \frac{\delta_{2a_2c_2} f_{2a_2c_2(t-1)k}}{F_{2k(t-1)}} (\mu_{1a_1c_1(t-1)h} + \mu_{2a_2c_2(t-1)k})$$

Finally, other assortative matings can be considered. Most often the structure of the parent population includes different groups involved in mating such as males on progeny testing, elite males or females, selected non elite males or females. For instance males in progeny testing may be mated exclusively to non elite females. To include assortative matings in the model, reproducers are classified according to mating groups m . $F_{msg(t-1)g}$ is the probability that a parent of group m , and sex s has genotype g . It is defined as:

$$F_{msg(t-1)} = \sum_{a,c} \delta_{msac} f_{sac(t-1)g}$$

where δ_{msac} is the proportion of animals of age a , class c and sex s within the m^{th} group.

When mating groups are considered, the equations of genotype frequency and EBV mean are modified accordingly. For instance in genotypic random mating:

$$f_{s11tg} = \sum_{h,k} \tau_{ghk} \sum_{m_1} \sum_{m_2} F_{m_1 1h(t-1)} F_{m_2 2k(t-1)}$$

$$\mu_{s11tg} = \frac{1}{2f_{s11tg}} \sum_{h,k} \tau_{ghk} \sum_{m_1} \sum_{m_2} \sum_{a_1, c_1} \sum_{a_2, c_2} \frac{\delta_{m_1 1a_1c_1} f_{1a_1c_1(t-1)h}}{F_{m_1 1h(t-1)}} \frac{\delta_{m_2 2a_2c_2} f_{2a_2c_2(t-1)k}}{F_{m_2 2h(t-1)}} (\mu_{1a_1c_1(t-1)h} + \mu_{2a_2c_2(t-1)k})$$

Optimization

The deterministic model described above allows a global optimization of the selection scheme in order to maximize the frequency of the desired genotype for a monogenic trait while minimizing the loss of genetic progress on a polygenic trait. These objectives may be achieved by taking into account the whole population (male and female) or a part of the population (female or male) for a given age. The decrease in genetic progress for the polygenic trait may be considered as a penalty term which may be defined either as a strict or as a progressive penalty. Under a strict penalty, the objective function is put to 0 when the genetic progress is below a threshold. Under a progressive penalty, this term is proportional to the decrease in genetic progress.

The frequencies of the favorable genotype and the average EBV in the population may be considered during all periods of selection (*e.g.* in the use of discounting) or only on the last selection year.

The decision variables w_{sactg} and α_{hkt} , describing the evolution of the population, have to be optimized. In our approach, we did not calculate the derivatives of the objective function, differing from the approaches used by Dekkers and van Arendonk (1998) and Manfredi *et al* (1998). A Genetic Algorithm (GA) was used to find the optimum solutions (Goldberg, 1989). A GA is an optimization technique based on an evolution of a population of artificial individuals. Each individual is a point in the research space and is characterized by its “chromosome” C_k , which determines the individual fitness $h(C_k)$; $k = 1, \dots, n$; n is a population size. The evolution process consists of successive generations. At each generation, individuals with high fitness are selected; the “chromosomes” of selected individuals are recombined and subjected to small mutations. A GA implemented as in Carroll (2001) was used in our study.

APPLICATION

This model was applied to an example, where selection for Scrapie resistance, which is controlled by a single gene (the *PrP* gene), was introduced as an additional selection criterion in an already existing dairy sheep selection scheme. It is expected that genetic progress in the polygenic trait (*e.g.* milk production) will be reduced under this selection objective.

Scrapie is a spongiform encephalopathy in sheep similar to BSE (Bovine Spongiform Encephalopathy) in Bovine and CJD (Creutzfeldt - Jakob disease) in Humans. In sheep, a genetic variability of the resistance to scrapie has been demonstrated mainly determined by the PrP polymorphisms (Elsen *et al.*, 1998; Fabre *et al.*, 2006). Thus, selection on the PrP locus was proposed to improve Scrapie resistance in sheep populations. Equations corresponding to this example are given in the Appendix.

The population (Figure 3) was composed of animals, males and females, aged from 0 to 6 yr. Females were selected on their own performances (with a selection rate W_{221}) after their 1st lactation between 2 and 3 yr of age (Figure 3). The selected females were distributed in 2 categories, elite (W_{232}) and non elite ($1 - W_{231}$). Among the young males born from matings between elite reproducers, a fraction W_{111} was mated to non elite females and their EBV's for the polygenic trait were estimated after a progeny test. A fraction W_{131} of these males was selected 2 yr later. Among the selected males, a proportion W_{142} was mated to the elite females to produce young male candidates for selection in the next generation. The replacement of females resulted from mating

between non elite females and elite, non elite males or males in progeny testing; and between elite females and elite males.

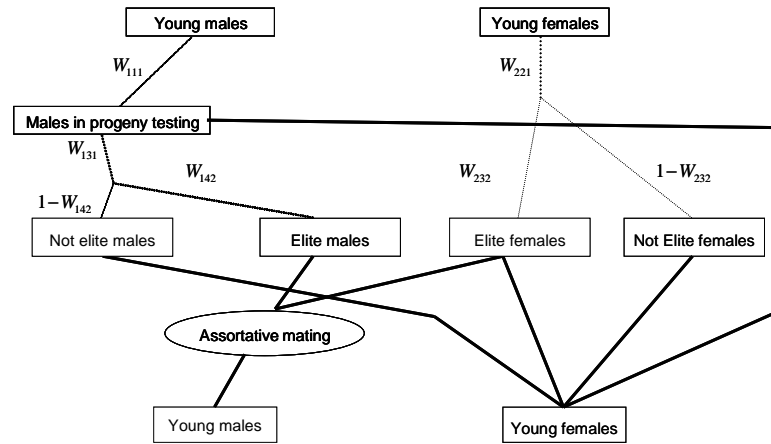


Figure 3. Population description with 7 categories (rectangles), 7 matings paths (thick lines), selection steps (dotted lines) and selection proportion W_{sac} (sex $s=1, 2$; age $a=1, \dots, a_s$; category $c=1, 2$)

2 alleles (favorable vs. unfavorable) were considered for the Prp gene, thus 3 genotypes exist ($g=1$ favorable genotype, $g=2$ intermediary genotype and $g=3$ unfavorable genotype). PrP genotypes were known for females and males. Only males were selected on their PrP genotype and on their EBV value before and after progeny testing. The proportion δ of males in progeny test among reproductive males was equal to 0.2. The proportion λ_a of females with age a (4, 5 and 6 yr) among reproductive females was 0.4, 0.3, and 0.3 respectively. This selection scheme was applied during 15 yr ($T=15$) with the following parameter values: $W_{221}=0.8$; $W_{232}=0.1$; $W_{111}=0.225$; $W_{131}=0.2$; $W_{142}=0.3$; $\sigma=36$. The accuracy for females (ρ_f) was 0.73. The accuracy for

males before and after progeny test was 0.56 (ρ_{m1}) and 0.77 (ρ_{m2}), respectively. The accuracy for elite males (ρ_{m3}) was 0.89. The GA parameters were: population size=100, number of generations=1,000, creep mutation probability=0.02, crossover probability=0.5, jump mutation probability=0.01.

The aim was to maximize Scrapie resistance, minimizing the loss of genetic progress in milk production. The objective function chosen was the frequency of homozygous favorable females ($s=2$, $g=1$) with a penalty on the genetic progress, constraining it not to be lower than a percentage π of the progress obtained without scrapie selection (ΔG_T). The objective function can be written as follows:

$$Fobj = \left(\frac{\sum_{a=1}^6 \sum_{c=1}^2 \frac{1}{a} f_{2acT1}}{\sum_{a=1}^6 \frac{1}{a}} \right) - 100 * \frac{\pi \Delta G_T - \sum_{g=1}^3 f_{211Tg} \mu_{211Tg}}{\pi \Delta G_T}$$

The 1st term is the mean of homozygous favorable females of the elite ($c=1$) and non elite ($c=2$) group weighted by the coefficient $\frac{1}{a}$ (a is the age) which gives more importance to younger females. All the females were considered in the objective function because all of them will participate in the renewal. The second term is proportional to the difference in genetic progress obtained in this selection scheme and without selection on Scrapie resistance.

First, the model was used with different initial frequencies of the favorable genotype without any constraint on genetic progress ($\pi=100\%$). 3 initial frequencies of the favorable genotype were considered (0.03, 0.3, and 0.69). Then, situations with different constraints on genetic progress were considered while fixing the initial frequencies. Finally, simplifications of the strategy were explored to make the optimized selection procedures more practicable for farms.

Effect of different initial frequencies without constraint on genetic progress

Whatever the initial frequency, the unfavorable genotype disappeared after 15 yr of selection (Table I).

Table 1. Final genotype frequencies for young females, objective function, and loss of genetic progress for 3 initial frequencies ($g=1$ favorable genotype, $g=2$ intermediary genotype, $g=3$ unfavorable genotype) without any constraint in the loss of genetic progress.

Genotype	Initial frequencies			Final frequencies for young females			Objective function	Loss of genetic progress (%)
	1	2	3	1	2	3		
	0.03	0.28	0.69	0.77	0.23	0.00	0.68	18.10
	0.30	0.50	0.20	0.91	0.09	0.00	0.88	6.40
	0.69	0.28	0.03	0.97	0.03	0.00	0.96	2.00

The loss of genetic progress was more important (18.1%) with a low initial frequency of the favorable genotype than with a high initial frequency (2.0%). This loss is much higher in the first few years, and impacts in the first few years are very important if counting discounted returns. The optimal selection proportions and mating rules evolve during the 15 yr process. The favorable genotype frequency increased, keeping a high selection proportion of the homozygous resistant males before and after their progeny test. When the initial frequency of the favorable genotype (0.03 or 0.3) was low, the selection proportions of males carrying this genotype were high (0.9 and 0.7) during the first years in order to rapidly fix this genotype in the sires. These selection proportions decreased progressively when all sires became resistant. When the initial frequency of the favorable genotype (0.69) was high, the selection proportions for the favorable genotype were lower, in particular before progeny testing (0.32 instead of 0.9 and 0.7). Moreover, the selection proportions after the progeny test were constant from the beginning, while they were constant only after the 9th and 4th yr for the case of low frequencies (0.03 and 0.3). Whatever the initial frequencies of the favorable genotype,

none of the susceptible males were selected, eliminating all of them from the beginning. A fully favorable population of sires was reached more rapidly with a high initial frequency of the favorable genotype (3rd yr) than with a low frequency (8th yr).

The reduction of the unfavorable genotype frequencies in the dam population occurred very slowly because (i) dams were selected just once, (ii) the selection threshold was the same across genotype classes and (iii) the selection pressure was low. With a high initial frequency of the favorable genotype, the objective was reached and the dam population was fully resistant. However with a low frequency, only sires were resistant at the end of the 15th yr, with the unfavorable genotype still segregating in the dam population.

Effects of different constraints on the loss of genetic progress

In contrast to the situation described above where no constraint was put on genetic progress, a 10% maximum loss of genetic progress at the end of the process was allowed ($\pi=90\%$) for the situation with the lower initial frequency (0.03). The case with the initial low frequency of the favorable genotype was studied because the loss of genetic progress with this frequency was the most important (18.1%). In this situation, the optimal selection proportions of males with a favorable genotype were lower and fluctuated between high and low values (Figure 4). These fluctuations can be explained by the fact that the genetic progress must be kept over a threshold. Each year, the selection intensity varies in order to compensate for the loss in genetic progress. Contrary to the previous case, the selection proportions of favorable and heterozygous genotypes were high and similar (0.7). The selection proportions for homozygous unfavorable sires were null. After the elimination of unfavorable genotype carriers, the

selection proportions of the favorable genotype (20%-30%) were higher than for the heterozygous genotype (5%-20%). The heterozygous genotype was conserved for a longer time than without any constraint on the genetic progress. The objective of achieving a fully resistant population of sires was reached less rapidly with 10% of maximum loss of genetic progress (11th yr) than without any constraint on genetic progress (8th yr). In the dam population, proportions of favorable homozygous and heterozygous genotypes were similar. Dams with an unfavorable genotype were almost entirely eliminated, being kept at less than 5%. Thus the optimal strategy of selection depends on the constraints put on genetic progress. If the priority is to reach a population fully resistant to the disease, the loss of genetic progress may not be essential, the selection on genotype information is drastic at the beginning of the process and the animals are quickly fully resistant. If a more important penalty is put on the loss of genetic progress, the selection changes, with cycles of high and low selection proportions for the favorable genotype. Eventually, the unfavorable genotype disappears and the population is divided between favorable homozygous and heterozygous genotypes.

In the previous case (Figure 4a), the selection proportions decreased gradually, as the population became progressively resistant. With a constraint on genetic progress, the fluctuation of selection proportions was larger (Figure 4b).

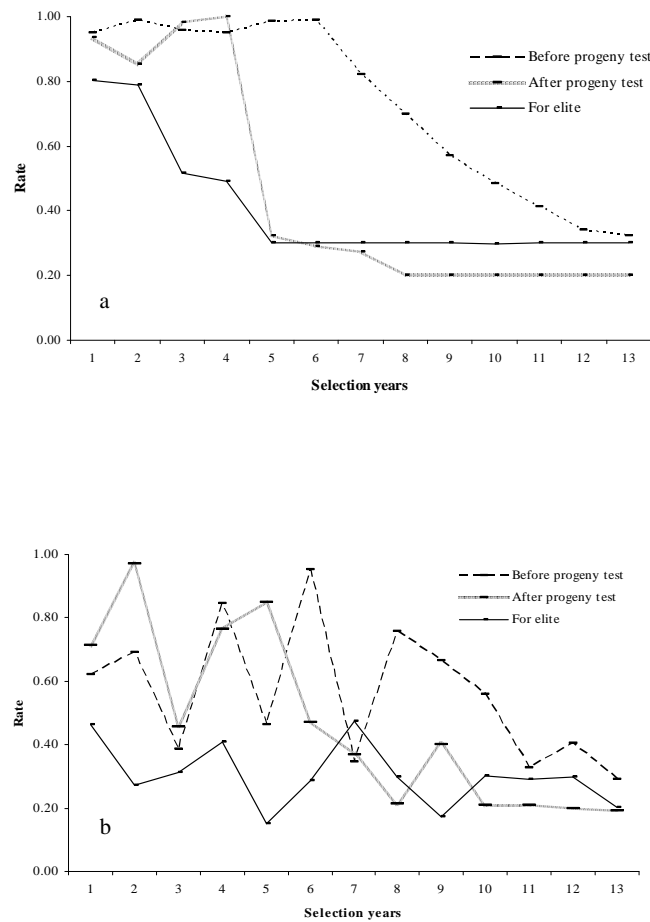


Figure 4. Selection rate evolution for the favorable genotype ($g=1$) without (a) and with (b) constraint in loss of genetic progress before progeny test (---), after progeny test (....) and for elite males (—)

The selection proportions of the favorable genotype were high during the first years, inducing an increase in resistance during this period; they were lower than with previous cases to maintain an acceptable genetic progress. This optimal strategy would

not be practicable for breeders owing to the annual variations of the selection proportions. However, a detailed analysis of selection proportions and assortative mating during the 15 selection yr, suggests some simplifications.

Simplifications of the optimal strategy

2 cases were analyzed: (1) a low initial frequency of the favorable genotype and a maximum 10% loss of genetic progress accepted after 15 selection years and (2) a low initial frequency of the favorable genotype without any constraint on the loss of genetic progress. The results are given in Table II. The first case shows the evolution of the objective function without any penalty on loss of genetic progress. The second case shows the evolution of the objective function with a maximum loss of 10 % in the genetic progress.

In the optimal strategy, the selection proportions of the males before progeny testing (W_{III} , Figure 3) varied between years and thus needed to be optimized. On the contrary, some selection proportions were constant after a few years of selection, suggesting some simplifications.

1. Within genotype selection proportions of males after progeny testing: The stabilization of these selection proportions is late, after the 11th selection year. However, the fixation of these selection proportions was tested. The values of the objective function were lower than in the optimal strategy. Thus, these selection proportions have to be optimized.
2. Within genotype selection proportions for elite males: The results of the previous optimization suggest a suboptimal strategy where these selection proportions are constant. The objective function value obtained with this simplification was similar

to the optimal one (Table II). Consequently, it appears that it is not critical for within genotype selection proportions for elite males to be optimised every year.

Table 2. Value of the objective function with some simplifications in 2 cases without constraint on loss of genetic progress (L-0) and with 10% of loss of genetic progress (L-10)

	L-0	L-10
ully optimized scheme	0.68	0.53
onstant selection rate for elite males	0.67	0.53
onstant selection rate after progeny testing	0.56	0.48
andom mating	0.68	0.52
onstant selection rate for elite males + random mating	0.67	0.52

Another question was to check if assortative mating based on the genotypes was necessary or not. The slight difference in the objective function value between the fully optimized scheme and the simplified one shows that mating optimization was not necessary.

Globally, 2 simplifications were thus possible: fixed selection proportions for elite males, and random mating. The case combining simplifications gave a value of the objective function similar to the optimal one. Moreover, the penalty weight of the loss on genetic progress was identical with or without simplifications. Implementing other simplifications could give a lower value of the objective function than the one obtained in the fully optimized scheme, but these were not considered here.

In our example, we proposed that the objective was to maximize the genotypic frequency of the favorable female genotypes with a penalty on genetic progress. Avoiding a loss of genetic progress increases the time needed to reach the fixation of the favorable genotype. For example, sires are all resistant at the end of the 8th yr or after 11th yr without any constraint or with a maximum loss of 10% of genetic progress, respectively. Other objective functions are possible and must be adapted to the aim of

the breeder. For example the males' frequency may be considered instead of the females' frequencies. Males have a more important impact on the selection scheme. In this example, the objective was to keep the homozygous genotype, however in the case of breeds with a risk of extinction it could be preferable to keep heterozygote genotypes. Another option could be to consider, the whole process rather than only the final state of the population, possibly using a discounting factor if sooner is better.

DISCUSSION- CONCLUSION

In this paper, a method was developed to optimize selection on an identified QTL or gene and a quantitative trait over multiple generations. The breeding program modeled assumed assortative mating and overlapping generations. This modeling may easily be adapted to a variety of situations concerning the underlying genetic model and the selection scheme. In this study, 2 alleles were considered but the methodology can consider a multi-allelic case. However, the number of alleles (and genotypes) should be limited. The number of optimized parameters increases in an exponential way with the number of alleles considered. For instance, when 3 genotypes (2 alleles) are considered, 10 parameters have to be optimized and, if 10 genotypes (4 alleles) are considered, the optimization problem would involve 108 parameters. Only the case with an additive QTL or gene was considered here, but dominance could be included. The method can also be extended to multiple QTLs or genes with any type of linkage between them (epistasis).

Different methodologies to solve multiple-generation selection problems in animal breeding are possible. In this study, a GA was used to solve the optimization

problem. Genetic Algorithms are more flexible when additional constraints have to be added than Optimal Control Theory, the method used by Dekkers and van Arendonk (1998), because in order to optimize selection the derivative of the selection criterion has to be calculated with Optimal Control Theory. Use of a GA by-passes this difficulty.

Additive variance was assumed to be constant over generations, and genotyping errors and relationship errors were assumed to be non-existent. The reduction of the additive variance by selection could be considered in the model.

The strategy developed here consisted of selecting animals that were resistant to a particular disease in order to decrease the number of infected animals. The aim was to reduce the incidence of the disease keeping the genetic progress on the traditional production traits.

In this model, inbreeding was not considered. A possibility for taking this into account would be to add the notion of family for males. Males of one family participate for the males' renewal for their family and males of different families participate for the females' renewal. This possibility would allow a certain genetic variability to be maintained.

Domestic animal diseases might represent a risk to human health. Thus, the identification of genes involved in disease resistance is an important genetic research subject. The scrapie case is an example where a national selection program for scrapie resistance was implemented as a precautionary measure. This dynamic and deterministic model was applied here in the scrapie case, however, it can be extended to other genes and selection schemes. For example, genes affecting important performance traits such as adiposity or meat quality in pigs (Le Roy *et al.*, 1990; Larzul *et al.*,

1997a; Hamilton *et al.*, 2000) or prolificacy in sheep (Davis, 2005; Fabre *et al.*, 2006) could use this model. Moreover, this deterministic model could evolve to a stochastic model for a limited number of breeding animals in a small population. Our model would need to be simplified for use in stochastic simulation, as stochastic simulations with a large number of choice variables can be impractical.

LITERATURE CITED

- Carroll, D. L. 2001. The genetic algorithm (ga) driver. <http://cuaerospace.com/carroll/ga.html>. Accessed Aug. 25, 2008
- Chakraborty, R., L. Moreau, and J. C. M. Dekkers. 2002. A method to optimize selection on multiple identified quantitative trait loci. *Genet. Sel. Evol.* 34:145-170.
- Davis, G. H. 2005. Major genes affecting ovulation rate in sheep. *Genet. Sel. Evol.* 37 Suppl 1:S11-23.
- Dekkers, J. C. M., and R. Chakraborty. 2001. Potential gain from optimizing multigeneration selection on an identified quantitative trait locus. *J. Anim. Sci.* 79:2975-2990.
- Dekkers, J. C. M., and J. A. M. van Arendonk. 1998. Optimizing selection for quantitative traits with information on an identified locus in outbred populations. *Genet. Resour. Crop. Ev.* 71:257-275.
- Elsen, J. M., J. Vu Tien Khang, F. Schelcher, Y. Amigues, J. P. Poivey, F. Eychenne, F. Lantier, and J. L. Laplanche. 1998. A scrapie epidemic in a closed flock of Romanov. *Proc. 6th world congress on genetics applied to livestock production, Armidale, Australia.* 27:269-272.
- Fabre, S., A. Pierre, P. Mulsant, L. Bodin, E. Di Pasquale, L. Persani, P. Monget, and D. Monniaux. 2006. Regulation of ovulation rate in mammals: Contribution of sheep genetic models. *Reprod. Biol. Endocrinol.* 4:20.
- Gibson, J. P. 1994. Short term gain at the expense of long term response with selection of identified loci. *Proc. 5th world congress of genetics applied to livestock production. University of Guelph, Ontario, Canada,* 21:201-204.
- Goldberg, D. E. 1989. *Genetics algorithms in search, optimization and machine learning.* ed. Addison-Wesley Publishing Company, Inc.
- Hamilton, D. N., M. Ellis, K. D. Miller, F. K. McKeith, and D. F. Parrett. 2000. The effect of the Halothane and Rendement Napole genes on carcass and meat quality characteristics of pigs. *J. Anim. Sci.* 78:2862-2867.
- Holland, J. H. 1973. Genetic algorithms and the optimal allocation of trials. *SIAM J. Comput.* 220:88-105.

- Lande, R., and R. Thompson. 1990. Efficiency of marker assisted selection in the improvement of quantitative traits. *Genetics* 124:743-756.
- Larzul, C., L. Lefaucheur, P. Ecolan, J. Gogue, A. Talmant, P. Sellier, P. Le Roy, and G. Monin. 1997a. Phenotypic and genetic parameters for longissimus muscle fiber characteristics in relation to growth, carcass, and meat quality traits in Large White pigs. *J. Anim. Sci.* 75:3126-3137.
- Larzul, R., E. Manfredi, and J. M. Elsen. 1997b. Potential gain from including major gene information in breeding value estimation. *Genet. Sel. Evol.* 29:161-184.
- Le Roy, P., J. Naveau, J. M. Elsen, and P. Sellier. 1990. Evidence for a new major gene influencing meat quality in pigs. *Genet. Res.* 55:33-40.
- Lewis, F. L. 1986. *Optimal control*. Wiley, New York.
- Manfredi, E., M. Barbieri, F. Fournet, and J. M. Elsen. 1998. A dynamic deterministic model to evaluate breeding strategies under mixed inheritance. *Genet. Sel. Evol.* 30:127-148.
- Meuwissen, T. H., and J. A. van Arendonk. 1992. Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. *J. Dairy Sci.* 75:1651-1659.
- Mongeau, M. 2003. *Exploitation de la structure en optimisation globale*, Laboratoire Mathématiques pour l'Industrie et la Physique. Handbook. UMR 5640. Toulouse, France
- Pong-Wong, R., and J. A. Woolliams. 1998. Response to mass selection when an identified major gene is segregating. *Genet. Sel. Evol.* 30:313-337.

APPENDIX: Equations of the example

2 yr females:

$$f_{221g} = f_{211(t-1)g}$$

$$\mu_{221g} = \mu_{211(t-1)g}$$

3 yr non-elite females:

$$f_{232tg} = f_{221(t-1)g} \frac{w_{221tg} (1 - w_{232tg})}{W_{221} (1 - W_{232})}$$

$$\mu_{232tg} = \mu_{221(t-1)g} + \frac{i(w_{221tg}) - w_{232tg}}{(1 - w_{232tg})} \frac{i(w_{221tg} w_{232tg})}{\rho_f \sigma}$$

where i() is the selection intensity.

3 yr elite females:

$$f_{231tg} = f_{221(t-1)g} \frac{w_{221tg} w_{232tg}}{W_{221} W_{232}}$$

$$\mu_{231tg} = \mu_{221(t-1)g} + \frac{i(w_{221tg}) - w_{232tg}}{w_{232tg}} \frac{i(w_{221tg} w_{232tg})}{\rho_f \sigma}$$

4 to 6 yr females

$$\mu_{2a1g} = \mu_{2(a-1)l(t-1)g}$$

$$f_{2a1g} = f_{2(a-1)l(t-1)g}$$

$$\mu_{2a2g} = \mu_{2(a-1)2(t-1)g}$$

$$f_{2a2g} = f_{2(a-1)2(t-1)g}$$

2 yr males before progeny testing

$$f_{121tg} = f_{111(t-1)g} \frac{w_{111tg}}{W_{111}}$$

$$\mu_{121g} = \mu_{111(t-1)g} + i(w_{111tg}) \rho_{m1} \sigma$$

2 yr males after progeny testing

$$f_{131tg} = f_{121(t-1)g}$$

$$\mu_{131tg} = \mu_{121(t-1)g}$$

4 yr non-elite males

$$f_{142tg} = f_{131(t-1)g} \frac{w_{131tg} w_{142tg}}{W_{131} W_{142}}$$

$$\mu_{142tg} = \mu_{131(t-1)g} + i(w_{131tg} w_{142tg}) \rho_{m3} \sigma$$

4 yr elite males

$$f_{141tg} = f_{131(t-1)g} \frac{w_{131tg} (1 - w_{142tg})}{W_{131} (1 - W_{142})}$$

$$\mu_{141tg} = \mu_{131(t-1)g} + \frac{i(w_{131tg}) - w_{142tg} i(w_{131tg} w_{142tg})}{1 - w_{142tg}} \rho_{m2} \sigma$$

Renewal

Newborn males

$$f_{111(t+1)g} = \sum_{h=1}^3 \sum_{k=1}^3 \tau_{ghk} \alpha_{hkt}$$

$$\mu_{111(t+1)g} = \frac{1}{2} \frac{1}{f_{111(t+1)g}} \sum_{h=1}^3 \sum_{k=1}^3 \tau_{ghk} \alpha_{hkt} \sum_{a_2=3}^6 \frac{f_{2a_2 2tk} \lambda_{a_2}}{\sum_{a_2=3}^6 f_{2a_2 2tk} \lambda_{a_2}} (\mu_{142th} + \mu_{2a_2 2tk})$$

Newborn females

$$f_{211(t+1)g} = \sum_{h=1}^3 \sum_{k=1}^3 \tau_{ghk} (\alpha_{hkt} W_{232} + \sum_{a_2=3}^6 \frac{f_{2a_2 1tk} \lambda_{a_2}}{(1-\delta)m_e W_{142} + (1-\delta)m_{ne} (1 - W_{142}) + \delta m_T})$$

$$\left[f_{142th} \left((1-\delta)m_e W_{142} - W_{f2} \left((1-\delta)m_e W_{142} + (1-\delta)m_{ne} (1 - W_{142}) + \delta m_T \right) \right) + f_{141th} (1-\delta)m_{ne} (1 - W_{142}) + \delta m_T f_{121th} \right]$$

$$\begin{aligned}
\mu_{211(i+1)g} &= \frac{1}{2} \frac{1}{f_{211(i+1)g}} \sum_{h=1}^3 \sum_{k=1}^3 \tau_{ghk} (\alpha_{hkt} W_{232} \sum_{a_2=3}^6 \frac{f_{2a_2 2tk} \lambda_{a_2}}{\sum_{a_2=3}^6 f_{2a_2 2tk} \lambda_{a_2}} (\mu_{142th} + \mu_{2a_2 2tk})) \\
&\quad + \sum_{a_2=3}^6 f_{142th} f_{2a_2 1tk} \lambda_{a_2} \left(\frac{(1-\delta)m_e W_{142}}{(1-\delta)m_e W_{142} + (1-\delta)m_{ne}(1-W_{142}) + \delta m_T} - W_{232} \right) (\mu_{142th} + \mu_{2a_2 1tk}) \\
&\quad + \sum_{a_2=3}^6 f_{141th} f_{2a_2 1tk} \lambda_{a_2} \left(\frac{(1-\delta)m_{ne}(1-W_{142})}{(1-\delta)m_e W_{142} + (1-\delta)m_{ne}(1-W_{142}) + \delta m_T} \right) (\mu_{141th} + \mu_{2a_2 1tk}) \\
&\quad + \sum_{a_2=3}^6 f_{2a_2 1tk} \lambda_{a_2} f_{121th} \left(\frac{\delta m_T}{(1-\delta)m_e W_{142} + (1-\delta)m_{ne}(1-W_{142}) + \delta m_T} \right) (\mu_{121th} + \mu_{2a_2 1tk})
\end{aligned}$$

Troisième partie

Modèle stochastique

Chapitre 5

Modèle stochastique

5.1 Résumé

La prise en compte de fluctuations de la réponse à la sélection dues à des effectifs limités est particulièrement importante à considérer quand la sélection se fait sur un gène identifié. En effet, si la fréquence initiale de l'allèle ayant un effet positif sur le caractère d'intérêt est faible, cet allèle peut disparaître par dérive. L'optimisation d'un plan de sélection assisté par gène, quand le modèle prend en compte ces fluctuations, pose néanmoins certains problèmes spécifiques.

Dans le cas stochastique, des décisions sont prises sans connaître exactement leurs conséquences alors que dans le cas déterministe, l'évolution du système et les apports associés sont parfaitement maîtrisés.

Les relations entre les différentes composantes du modèle (paramètres, variables de décision et d'état) ne correspondent plus à des statistiques classiques mais peuvent être basées sur des statistiques d'ordre. L'augmentation du nombre de classes et d'étapes de sélection entraîne une modélisation très complexe.

Dans le modèle déterministe précédent, les variables de décision à optimiser sont définies intra classe et par année. Avec cette modélisation, le nombre de variables à optimiser est important et peu compatible avec une approche stochastique qui est beaucoup plus coûteuse en temps de calcul. Nous nous sommes donc focalisés sur des modalités simplifiées de sélection répétées identiquement au cours du temps, ou linéairement dépendante de l'état de la population à l'instant t , définie par les fréquences des génotypes en cet instant.

Cette approche n'efface pas la difficulté due à la non unicité de la réponse à une décision. A un jeu de valeurs des variables décisionnelles ne correspond pas une valeur unique de la fonction objective mais un ensemble de valeurs. En fait, une décision à l'instant t n'est optimale que si on prend en considération l'ensemble des événements à venir affectés de leur probabilité de réalisation qui dépendent des décisions à optimiser pour l'avenir ($t + 1$, $t + 2$, ..., T). L'utilisation de simulations avec un modèle dont les parties élémentaires sont les individus permet d'obtenir aisément des informations sur la variabilité de la réponse.

Notre approche considère une population subdivisée en éléments simples : individus définis par leur sexe, âge, catégorie (élite/non élite). Chaque individu est caractérisé par son génotype et sa valeur polygénique. L'optimisation de la sélection se fait sur des pénalités que l'on ajoute aux estimations des valeurs polygéniques. Ces pénalités varient en fonction du génotype au gène identifié et éventuellement en fonction des fréquences génotypiques au sein de la population au moment de l'application de la pénalité. Pour un jeu de valeurs de ces pénalités, l'évolution de la population est simulée plusieurs fois afin d'obtenir la distribution des variables d'état. La fonction objective est une statistique (espérance) des distributions des génotypes du gène affectant le caractère secondaire et des EBV du caractère principal à la fin du processus de sélection.

Notre objectif a été de montrer, sur un exemple simple, que la démarche était possible et efficace. Nous avons également montré comment elle pouvait être utilisée pour faire des choix de méthodes de sélection. Pour cela nous avons utilisé, comme support, un modèle stochastique dérivé du modèle déterministe décrit dans le chapitre précédent.

Dans la plupart des populations d'animaux de ferme seuls les effectifs de mâles sont limités. Dans notre modèle, les éléments de la population sont les individus pour les mâles, et des classes pour les femelles. Les variables d'état caractérisant ces éléments sont leur génotype et leur valeur polygénique pour chaque mâle, la fréquence génotypique et la valeur polygénique moyenne pour chaque classe de femelles. Pour faciliter la démonstration, le schéma de sélection a été simplifié à l'extrême : pas de sélection de la voie femelle ; sélection en une seule étape des mâles (N_s retenus pour N candidats), accouplements au hasard, générations séparées, variance génétique constante au cours du temps. L'optimisation sous contrainte de la fonction objective a été obtenue à l'aide de l'algorithme génétique.

Nos résultats concernant la comparaison entre le modèle déterministe et le modèle stochastique décrivant le même type de population a montré l'utilité d'une approche stochastique. La stratégie de sélection optimale ne peut pas être réduite à celle obtenue avec un modèle déterministe.

D'autre part, les résultats obtenus dans le cadre de cette modélisation stochastique montrent qu'un critère de décision très « simple » donne d'aussi bons résultats qu'un critère de sélection prenant en compte la structure génotypique de la population (fréquence génotypique des femelles ou des mâles et des femelles). Cependant l'application considérée est tellement simplifiée (générations séparées et sélection sur les mâles uniquement et en une étape) que les résultats obtenus ne sauraient être généralisés. De plus, nous nous sommes restreints dans cette application à l'optimisation de l'espérance de la fonction objective. D'autres choix auraient pu être faits comme par exemple l'optimisation de l'espérance de la fonction objective avec une variance minimale.

Cette première approche a permis de répondre aux questions concernant l'utilité et la faisabilité de considérer un modèle stochastique. Néanmoins, il serait intéressant de mettre en oeuvre des situations plus complexes. Par exemple, les « meilleurs » mâles pourraient être conservés sur plusieurs générations, plusieurs étapes de sélection pour les femelles comme pour les mâles pourraient être considérées. Il serait également intéressant de prendre en compte l'évolution de la variance génétique au cours des générations. Enfin, dans le contexte qui nous

préoccupe d'un nombre limité de reproducteurs, la prise en considération de la matrice de parenté entre les individus pourrait aider à minimiser le taux de consanguinité et maintenir une certaine diversité génétique.

5.2 Article

Cet article va être soumis à la revue « Journal of Animal Science ».

1

2 **A stochastic model to optimize a gene assisted selection in a population**
3 **of small size¹**

4

5 A.D. Costard*, J.-M. Elsen*

6

7 *INRA, SAGA station d'amélioration génétique animale, 31326 Castanet-
8 Tolosan, France

9

10

11

12 **Running head:** optimizing selection in a small population

13

14

15

16

17

18

19

20

21

22

23 ¹Corresponding author: Anne.Devalle@toulouse.inra.fr

ABSTRACT

Many of the models used to optimise selection process in livestock make the assumption that the population is of infinite size. The finite size case should however be considered when selection involves one identified gene. Indeed, it is likely that a favourable allele of this gene may disappear through drift if its initial frequency is low. In this paper, a stochastic model was developed to simultaneously optimize selection on two traits in a limited size population: a quantitative trait with underlying polygenic variation and a monogenic trait. We outline the interests of considering the limited size of the population in a stochastic modeling through a simple example. Such stochastic models pose some technical problems (convergence to the maximum, computational burden) which could obliterate their usefulness as compared to simpler but approximate deterministic models where the population is supposed to be of infinite size. Through a simple example, we show the feasibility of the optimization of this type of model with the genetic algorithm and demonstrate its interest relatively to the corresponding determinist model which assumes that the population is of infinite size.

Key words: Gene Assisted Selection, Optimization, Small Population, Genetic Algorithm, Stochastic Model

INTRODUCTION

Many studies have demonstrated that marker-assisted selection (MAS/GAS) is more efficient than classical selection schemes for traits with low heritability or difficult and/or expensive to measure (Dekkers and van Arendonk, 1998; Lande and Thompson, 1990; Meuwissen and van Arendonk, 1992). The superiority of MAS has been demonstrated on the short term (Gibson, 1994; Larzul *et al.*, 1997; Meuwissen and Goddard, 1996; Pong-Wong and Woolliams, 1998; Spelman and van Arendonk, 1997) but these results were not confirmed on the long term (Gibson, 1994; Verrier, 2001). Several procedures to optimise selection with overlapping generations have been proposed, either to maximise the response to the selection with an optimal control approach (Chakraborty *et al.*, 2002; Dekkers and Chakraborty, 2001) or with sequential quadratic programming (SQP, (Manfredi *et al.*, 1998)). Other studies also attempted to maximise the genetic progress for one or several traits while limiting the inbreeding coefficient to a maximum value (Colleau *et al.*, 2004). Costard *et al.* (2009) developed a mathematical approach to simultaneously optimise the selection on two traits: a quantitative trait with underlying polygenic variation and a monogenic trait. They proposed to maximise the frequency of the favourable genotype for the monogenic trait (*e.g.* resistance to a disease) while minimising the loss of genetic progress on the polygenic trait. The models that are used to optimise selection process in livestock usually

69 make the assumption that the population is of infinite size. Few studies took
70 into account the variability in the response due to finite size, and mostly in
71 the classical selection scheme situation (Meuwissen, 1997; Pong-Wong and
72 Woolliams, 2007; Villanueva *et al.*, 2006). The finite size case should
73 however be considered when selection mostly focus on one identified gene.
74 Indeed, it is likely that a favourable allele for the desired genotype may
75 disappear through drift if its initial frequency is low.

76 In this study, we address the question of the optimisation of a marker-
77 assisted selection scheme that takes into account the variations due to finite
78 population size. To this aim, we have further improved the deterministic
79 model of Costard *et al* (2009) with a stochastic model and, in this paper, we
80 demonstrate the feasibility and the relevance of the proposed approach.

81

82 MATERIAL AND METHODS

83

84 *The deterministic approach*

85 The model proposed by Costard *et al* (2009) was developed to optimise (1)
86 the selection of candidates after their breeding value evaluation for a main
87 polygenic trait and for a secondary monogenic trait and (2) the mating
88 between selected individuals. In this study, we will only describe the
89 selection part of the model.

90 The deterministic model proposed by Costard *et al* (2009) can be described
91 by a set of equations that include fixed parameters, decision variables (*e.g.*
92 selection rate) and state variables. In the case of a deterministic model, these
93 latter variables are statistics which describe elements of the whole
94 population. These elements may be individuals or classes of individuals that
95 share the same characteristics across the population. In Costard *et al*
96 (2009), the elements are classes defined by the sex s , the age a , the genotype
97 g for a gene to be selected and the category c (*e.g.* elite vs. non elite for
98 mating males and females). The state of a class is described by its relative
99 frequency and by the expectation of the polygenic trait distribution. The
100 equations model the time evolution of the state variables with respect to
101 natural (ageing or death) and artificial (selection) phenomena with discrete
102 steps.

103 The vector of the state variables at time $t + 1$ depends on state variables, and
104 parameters values at time t as well as decision variables values
105 corresponding to the applied decision between times t and $t + 1$.
106 Furthermore, the breeding objective is described by an objective function
107 that depends on the parameters and the variables in the model. For example,
108 this function can be the favourable genotype frequency in the reproducers
109 that were born in the last selection year.

110 With this type of modelling, a set of parameters and decision variable values
111 corresponds to one and only one evolution of the state variables, and,

112 therefore, to only one value of the objective function. To maximise the
 113 objective function, the optimisation of the decision variables is performed
 114 using optimal control algorithms (Chakraborty *et al* 2002, Dekkers and
 115 Chakraborty 2001), sequential quadratic programming (SQP, Manfredi *et al*
 116 1998) or genetic algorithm (Costard *et al*, 2009). Note that the selection
 117 process being dynamic, the best decision at time $t+1$ depends on the state
 118 variables values at time t , and it is needed to simultaneously optimise the
 119 whole set of decision variables across time. In other words, the optimisation
 120 results in a set of decisions which evolve across time when the population
 121 state changes, especially for the genotype frequency on the identified gene.
 122 For example, it may be optimal to apply a high selection pressure on the
 123 candidates class which carry the favourable genotype at the beginning of the
 124 process, when the genotype is not frequent, and an medium selection
 125 pressure at the end of the process, when the genotypes becomes dominant.
 126 In practice, in Costard *et al* (2009) modelling, the decision variables are
 127 selection rates, on genetic values for the polygenic trait of individuals
 128 belonging to key classes (defined by some s , a , c) in the population. Note
 129 that mating probabilities between elite individuals according to their
 130 genotype can also be introduced in the model. If individuals belonging to a
 131 class C_{sagc} are involved in a selection step, the optimisation will provide the
 132 selection rates q_{sagcgt} on their estimated breeding value (EBV) that must be
 133 applied to candidates with genotype g at time t , given the constraint that the

134 mean selection rate across the three genotypes (for a biallelic gene), must be
 135 equal to a fixed q_{sac} . Other decision variables could have been used without
 136 affecting the meaning of the model: an EBV threshold value λ_{sacgt} above
 137 which the candidates would be selected, or the addition of a penalty term
 138 θ_{sacgt} to the EBV. The state variable ‘mean polygenic value’ $\mu_{sacgt+1}$ of the
 139 progeny at time $t+1$ resulting from the selection within C_{sacg} at time t is
 140 function of μ_{sacgt} , the relative genotypes frequencies f_{sacgt} and the rates,
 141 thresholds or penalty terms q_{sacgt} , λ_{sacgt} , θ_{sacgt} .

142

143 ***Proposed stochastic approach***

144 In the case where the population size is finite, one must take into account
 145 the variability of possible states at time $t+1$ resulting from a decision taken
 146 at time t . This variability generates two specific phenomena.

147

148 Firstly, the relationship between the parameters, decision and state variables
 149 – elementary statistics on EBV and genotypes distributions of the population
 150 *elements*, must be adapted, for instance using order statistics. However, the
 151 modelling problem quickly becomes inextricable when the number of
 152 classes and selection cycles increase. A simpler approach would be to
 153 consider as model *elements* the individuals, rather than their classes. For
 154 example, if the number of 4-years old elite males is limited to 5 individuals,

the 3 elements of the corresponding model (three classes defined by the possible genotypes on a biallelic gene) will be replaced by 5 elements (the 5 individuals sharing the same sex, age and category), but each possessing its own state defined by its genotype and polygenic value.

This approach does not solve the difficulties linked to the fact that there is not a unique response at time $t+1$ (described by the state values at this time), to a decision taken at time t , and, thus, that there is not a unique value of the objective function corresponding to a set of decision variables.

However, this type of modelling enables to easily simulate the selection process, to obtain a distribution of possible responses, and thus to estimate parameters of the objective function distribution, such as its expectation, for any set of decisions. Several criteria can be chosen as objective function, such as minimising the probability that the favourable genotype frequency, or a combination of favourable genotype frequency and genetic progress on the polygenic trait is below a fixed threshold.

It must be emphasized that, with this stochastic modelling approach, the decision variables (rates, thresholds or penalties q_{sacgt} , λ_{sacgt} ou θ_{sacgt}) are not equivalent. We do not think in terms of a proportion of selected individuals (q_{sac}) but in terms of number of selected candidates Ns amongst N individuals. In this case, only the penalty terms are easily applicable: the Ns candidates will be the ones with the largest penalised EBV.

176

177 Secondly, the search for the optimal selection process should take into
178 account the possible events to come: a decision taken at time t will be
179 optimal only if the set of following events weighted by their probabilities
180 are considered. These probabilities depend on the decisions to be optimised
181 at time $t+1, t+2, \dots, T$. Thus the rule used in the deterministic modelling
182 (simultaneous optimisation of the whole set of decision variables across
183 time in order to deal with the fact that the best decision at time $t+1$ depends
184 on the state variables values at time t) is not practically applicable. For
185 instance, dynamic programming, although often used in this framework,
186 cannot be applied in our particular case where the state variables are
187 numerous continuous variables instead of a few discrete variables.

188

189 In the deterministic models proposed by Chakraborty *et al* 2002, Dekkers
190 and Chakraborty (2001), Manfredi *et al* (1998) and Costard *et al* (2009),
191 optimal penalties (or equivalently to the selection rates) were determined for
192 each year and class (sex/genotype/category) or (sex/age/category/genotype)
193 I.e., the *elements* of the model. In the present stochastic modelling, we
194 propose to focus the optimisation on a simplified framework, with within
195 genotypes penalties constant across time, or with penalties linearly
196 dependent on the population state at time t , such as the genotype frequency
197 at time t .

198 To summarise, we propose the following approach: The population is
199 analysed with simple *elements*: the individuals characterized by their sex,
200 age, category (elite/non elite). Their state is defined by their genotype g and
201 their polygenic value m .

- 202 1. The optimisation of the selection consists in choosing the penalties
203 to be added to the estimated breeding values (EBV). These penalties
204 vary according to the genotype at the identified gene and possibly
205 according to the genotype frequencies when applying the penalty.
- 206 2. For a given set of penalties, the population evolution is simulated
207 many times so as to obtain the distribution of the state variables.
- 208 3. The objective function is a composite statistic of the distribution, at
209 some stage of the selection process (*e.g.* the last year) and across
210 elements of the models defined by some characteristics (*e.g.* the 1
211 year old females), of genotypes at the gene affecting the secondary
212 trait and of polygenic values for the main trait.

213 Besides the computational time inherent to any stochastic approach, the
214 proposed approach can be invalidated for two reasons:

- 215 1. All numerical optimisation procedures are based on the evaluation of
216 differences between objective function values obtained for decision
217 variables vector which are closer and closer to each other when the
218 optimum is nearly reached. It may be possible that the estimation
219 errors of the objective function are of the same order of these

220 variations, which would render the convergence towards an optimum
221 impossible.

222 2. We cannot exclude that the approximate solutions given by a simpler
223 deterministic model was sufficient if one is solely interested in the
224 distribution expectations.

225 In this study, our aim is to show that our proposed approach is feasible and
226 efficient on a simple example. We will then illustrate how this approach can
227 be used to choose amongst different selection methods.

228

229 ***Proof model***

230 In most livestock populations, only the males numbers are small. Our
231 illustrative model is therefore composite, as the population elements are
232 individuals for the males and classes for the females. Consequently, the state
233 variables which characterise these elements are the genotype g_j and the
234 polygenic value m_j for each male, the genotypic frequency $f_{s,a,c,t,g}$ and the
235 mean polygenic value $\mu_{s,a,c,t,g}$ for each class of females ($s=2$).

236 To facilitate the proof of concept, the selection scheme is extremely
237 simplified (Figure 1). There is no female selection, the male selection is
238 performed in one step (N_s selected candidates amongst N candidates), and
239 we include random mating, non-overlapping generations and constant
240 genetic variability across time. We use the notations of Costard *et al* (2009)

with the appropriate modifications to take into account the finite male population size.

Modelling the renewal

Females renewal. The genotypic frequency and the mean polygenic values of the new generation ($a=I$, $c=I$) at time t are defined as:

$$f_{211tg} = \frac{1}{Ns} \sum_{i=1}^{Ns} \sum_k \tau_{ghik} \eta_i F_{2(t-1)k}$$

$$\mu_{211tg} = \frac{1}{2Ns f_{211tg}} \sum_{i=1}^{Ns} \sum_k \tau_{ghik} \eta_i \sum_{a_2, c_2} \frac{\delta_{2a_2c_2} f_{2a_2c_2(t-1)k}}{F_{2(t-1)k}} (m_i + \mu_{2a_2c_2(t-1)k})$$

where $\eta_i = \begin{cases} 1 & \text{if the male } i \text{ is selected} \\ 0 & \text{otherwise} \end{cases}$

τ_{ghik} is the probability that the progeny has a genotype g given that the parents have the genotypes k (dam) and h_i (sire i) $F_{2(t-1)k}$ is the probability that the dam has a genotype k , defined by:

$$F_{2(t-1)k} = \sum_{a,c} \delta_{2ac} f_{2ac(t-1)k}$$

where δ_{2ac} is the proportion of females of age a and class c .

Males renewal. We randomly simulate the genotype $g_i=g$ and the polygenic value m_i of the sire i ($i=1 \dots N$), which is the progeny of the sire j of genotype h_j , and of polygenic value m_j . The genotype is simulated from a multinomial

258 distribution with parameters $p_g = \sum_k \tau_{gh,k} F_{2(t-1)k}$ ($g=1, 2, 3$). The polygenic
 259 value is sampled from a normal distribution with a constant variance and
 260 expectation $\frac{1}{2} \sum_k \tau_{gh,k} \sum_{a_2, c_2} \frac{\delta_{2a_2c_2} f_{2a_2c_2(t-1)k}}{F_{2(t-1)k}} (m_j + \mu_{2a_2c_2(t-1)k})$

261

262 ***Modelling the reproducers selection***

263 For the generation t , an index $I(j, t) = m_j + c(t, h_j)$ is assigned to each
 264 male candidate j . This index is a linear combination of its polygenic value
 265 m_j and a weight $c(t, h_j)$, which is function of its genotype h_j . The N_s
 266 males with the largest index values $I(j, t)$ are selected amongst the N
 267 candidates to renew the population.

268 The weight formulation $c(t, h_j)$ can be optimised. For example, a selection
 269 uniquely based on the main polygenic trait is obtained using the index
 270 $I(j, t) = m_j$. Conversely, the criterion $I(j, t) = c(t, h_j)$ corresponds to a
 271 selection that is only based on the secondary trait. Generally, the coefficients
 272 $c(t, h_j)$ can be constant or depend on generation t population characteristics
 273 such as males and/or females genotypic frequencies. The optimisation of the
 274 reproducers selection, and therefore of the selection process first consists in
 275 choosing the type of weights, and then in determining the best values to be
 276 assigned to decision variables for each type of weight. In our proof model,

three types of weight and indexes (I_1 , I_2 , I_3), which correspond to
 respectively 10, 6, and 2 decision variables are compared:

1. The weights $c(t, h_j)$ can be a linear combination of the males and
 females genotypic frequencies in the population at generation t :

$$\begin{cases} c(t, 1) = 0 \\ c(t, 2) = \theta_1 f_{2,1,1,t,1} + \theta_2 f_{2,1,1,t,2} + \theta_3 f_{1,1,1,t,1} + \theta_4 f_{1,1,1,t,2} + \theta_5 \\ c(t, 3) = \theta_6 f_{2,1,1,t,1} + \theta_7 f_{2,1,1,t,2} + \theta_8 f_{1,1,1,t,1} + \theta_9 f_{1,1,1,t,2} + \theta_{10} \end{cases}$$

The 10 parameters to be optimised are in this case $\theta_1, \dots, \theta_{10}$. The
 information about the males genotypic frequencies is neglected as they have
 a low economical impact for the production given their small population
 size.

286

$$\begin{cases} c(t, 1) = 0 \\ c(t, 2) = \theta_1 f_{2,1,1,t,1} + \theta_2 f_{2,1,1,t,2} + \theta_3 \\ c(t, 3) = \theta_4 f_{2,1,1,t,1} + \theta_5 f_{2,1,1,t,2} + \theta_6 \end{cases}$$

288

The 6 parameters to be optimised are θ_1 , θ_2 , θ_3 , θ_4 , θ_5 , θ_6 .

At last, the third type of weight is the simplest case where the coefficients
 $c(t, h_j)$ are constant for any generation t :

$$\begin{cases} c(t, 1) = 0 \\ c(t, 2) = \theta_1 \\ c(t, 3) = \theta_2 \end{cases}$$

293 In this particular case, there are only 2 parameters to optimise θ_1, θ_2 .

294

295 ***Objective function***

296 To facilitate the comparison between the stochastic and the deterministic
 297 models, the objective function to be maximised is the expectation of a
 298 composite function which describes the population state at the end of the
 299 selection process for the secondary trait (favourable female genotype
 300 frequency) and for the main trait (deviation from the maximally achievable
 301 genetic progress):

$$302 \quad \max E(F_{obj}) = \max E\left(f_{2,1,1,T,1} - \lambda \frac{\pi \Delta G - \Delta G'}{\pi \Delta G}\right)$$

303 where $\Delta G'$ is the genetic progress achieved on the main trait for the
 304 simultaneous selection on both traits, ΔG is the genetic progress that would
 305 have been achieved without selection on the secondary trait and π is the
 306 percentage of tolerated loss of genetic progress. Following Costard *et al*
 307 (2009), this objective function is therefore a combination of the female
 308 frequencies carrying the desired genotype at the last generation and a
 309 penalisation on the loss of genetic progress, according to a weight λ chosen
 310 by the breeder. As in Costard *et al* (2009), the optimisation under constraint
 311 of the objective function was obtained with a genetic algorithm.

312

313 *Numerical application*

314 The population consists of $N = 10$ males, amongst which the $N_s = 5$ best
 315 ranked males according to their index are selected. After this selection,
 316 males and females are mated in a panmictic manner, with an identical male
 317 contribution (one male gives 2 males progenies for each generation). This
 318 process is iterated on T generations (5, 10, 15, 20) and simulated 1000
 319 times. The 3 considered genotypes ($g=1,2,3$) have respective initial
 320 frequencies of 20%, 50%, 30%. The genetic variance is equal to $\sigma_g^2=36$. In
 321 this example, the coefficient λ is set to an arbitrary value of 5.

322

323 **RESULTS**

324 *Feasibility of the model*

325 We compared the results obtained by the genetic algorithm with those
 326 obtained by a grid search for the third index I_3 with two decision variables
 327 (θ_1, θ_2) . Three different cases were tested with 5, 10 and 15 generations with
 328 a maximum loss of progress of 10%. Both approaches gave similar results
 329 (Table 1). Results are better with genetic algorithm. These results could be
 330 explained by the imprecise step of the grid search (0.2) and by the limited
 331 number of simulation (100). Therefore, for this selection scheme, a global
 332 optimisation with a stochastic modelling approach seems feasible with a
 333 genetic algorithm.

334 ***Superiority of the stochastic model compared to the deterministic model***

335 The deterministic model of Costard *et al* (2009) has been simplified to fit
 336 into the stochastic model framework described above. For any generation,
 337 the male selection is performed with a constant rate W_{111} as in Costard et al
 338 (2009). The characteristics of the selected males, such as genotypic
 339 frequencies and polygenic values are obtained as follows:

$$f(1,2,2,t,g) = \frac{(1 - \phi(\frac{\beta_2 - \mu(1,1,1,t,g) - c(g,t)}{\sigma}))f(1,1,1,t,g)}{W_{111}}$$

$$\mu(1,2,2,t,g) = \mu(1,1,1,t,g) + i(1 - \phi(\frac{\beta_2 - \mu(1,1,1,t,g) - c(g,t)}{\sigma}))\sigma$$

341 The mating between the selected males and the females is randomly
 342 performed. The genotypic frequency and the mean polygenic value of the
 343 progeny are:

$$f(s,1,1,t+1,g) = \sum_h \sum_k \tau_{ghk} f(1,2,2,t,h) f(2,1,1,t,k)$$

$$\mu(s,1,1,t+1,g) = \frac{1}{2f(s,1,1,t+1,g)} \sum_h \sum_k \tau_{ghk} f(1,2,2,t,h) f(2,1,1,t,k) (\mu(1,2,2,t,h) + \mu(2,1,1,t,k))$$

347 Table 2 displays the optimisation of the deterministic model given the
 348 different configurations: 5, 10, 15 and 20 generations, 10% loss of genetic
 349 progress or no constraint on the loss. The values of the objective function
 350 obtained with the deterministic model are in general greater or equal to
 351 those obtained with the stochastic model. These results were logical as the

352 deterministic model did not consider the depressing effect of the random
353 sampling on the mean distributions. The finite population size therefore has
354 a noticeable effect on the selection outcome. The optimal values of the
355 prediction of the coefficients θ_1, θ_2 that were obtained without constraint on
356 the loss of genetic progress were very high for both stochastic and
357 deterministic model. However, these values strongly decrease when a
358 constraint on the loss of genetic progress is imposed. In addition, with this
359 constraint, the values of the coefficients θ_1, θ_2 were very close for both
360 models.

361 In order to assess the difference in prediction between the two models,
362 simulations were performed with the stochastic model using values of the
363 coefficients θ_1, θ_2 obtained with the optimisation of the stochastic model
364 (“stochastic coefficients”) and values obtained with the optimisation of the
365 deterministic model (“deterministic coefficients”) (Table 3). In the case
366 where no constraint were imposed on the genetic progress, the use of a
367 stochastic model is not justified as the results obtained with both models
368 were identical.

369 In the case where a constraint was imposed, there was a significant
370 difference between outputs of the stochastic model using stochastic or
371 deterministic coefficients ($p < 0.001$): the objective function value of the
372 stochastic model was smaller using the deterministic rather than the

373 stochastic coefficients whatever the generations number, with a lower loss
 374 of genetic progress is less important and lower female genotypic frequency .
 375 As our aim is to obtain a maximum genotypic frequency, a stochastic model
 376 rather than a deterministic model seems therefore appropriate.

377

378 ***Application of the stochastic model to choose a selection criterion***

379 The stochastic model was evaluated with different indexes I_1, I_2, I_3 on 5,
 380 10, 15 and 20 generations, with different values of accepted genetic progress
 381 loss: 5%, 10% and 100%. Table 4 shows that the values of the optimised
 382 objective function were similar for any type of index. Therefore, taking into
 383 account the population structure with I_1, I_2 does not seem useful. In our
 384 setting with non-overlapping generations and a selection of the males only
 385 in one step, the I_3 index is sufficient.

386 Regarding the coefficient values θ_1, θ_2 for the index I_3 and the different
 387 tested configurations, the coefficients assigned to the males of genotype 2
 388 and 3 were negative (Table 2). The coefficient assigned to the genotype 2
 389 was less penalising than the one assigned to the genotype 3.

DISCUSSION- CONCLUSION

The feasibility of the optimisation with a stochastic modelling using a genetic algorithm was validated through a grid search in a very simple framework.

The comparison made between the deterministic and stochastic models demonstrated the usefulness of a stochastic approach: an optimal selection strategy obtained with this approach could not be obtained with a deterministic model.

In a stochastic modelling framework, the results obtained with a 'simple' decision criterion were as good as those obtained with a selection criterion that takes into account the genotypic structure of the population, such as females and/or males genotypic frequency. However, our framework was restricted to non overlapping generations and the selection of the males only in one step. The generalisation of the obtained results should therefore be taken with caution.

It would be interesting to set into more complex frameworks. For example, the 'best' males could be kept during several generations, or several selection steps for both females and males could also be considered. It would also be interesting to take into account the evolution of the genetic variance across generations. Finally, in the specific setting of a finite population size of reproducers, considering the relationship matrix between the individuals may help minimising the inbreeding rate and maintaining a

412 genetic diversity.

413 To illustrate the application of the proposed stochastic approach, we have
414 restrained our study to the optimisation of the objective function
415 expectation. Other criteria should have been used, such as the maximisation
416 of the objective function expectation with a constraint on the variance.

LITERATURE CITED

- 417
418
419
420 Chakraborty, R., L. Moreau, and J. C. M. Dekkers. 2002. A method to
421 optimize selection on multiple identified quantitative trait loci. *Genet.*
422 *Sel. Evol.* 34: 145-170.
- 423 Colleau, J.-J., S. Moureaux, M. Briend, and J. Bechu. 2004. A method for
424 the dynamic management of genetic variability in dairy cattle. *Genet. Sel.*
425 *Evol.* 36: 373-394.
- 426 Costard, A. D., Z. G. Vitezica, C. R. Moreno, and J. M. Elsen. 2009. A
427 dynamic deterministic model to optimize a multiple-trait selection
428 scheme. *J. Anim Sci.* 87: 885-894.
- 429 Dekkers, J. C. M., and R. Chakraborty. 2001. Potential gain from
430 optimizing multigeneration selection on an identified quantitative trait
431 locus. *J. Anim Sci.* 79: 2975-2990.
- 432 Dekkers, J. C. M., and J. A. M. van Arendonk. 1998. Optimizing
433 selection for quantitative traits with information on an identified locus in
434 outbred populations. *Genet. Resour. Crop. Ev.* 71: 257-275.
- 435 Gibson, J. P. 1994. Short term gain at the expense of long term response
436 with selection of identified loci. In: 5th World Congr. Of Genet. Appl.
437 *Livest. Prod.* P 201-204.
- 438 Lande, R., and R. Thompson. 1990. Efficiency marker assisted selection

- 439 in the improvement of quantitative traits. *Genetics* 124: 743-756.
- 440 Larzul, R., E. Manfredi, and J. M. Elsen. 1997. Potential gain from
441 including major gene information in breeding value estimation. *Genet.*
442 *Sel. Evol.* 29: 161-184.
- 443 Manfredi, E., M. Barbieri, F. Fournet, and J. M. Elsen. 1998. A dynamic
444 deterministic model to evaluate breeding strategies under mixed
445 inheritance. *Genet. Sel. Evol.* 30: 127-148.
- 446 Meuwissen, T. H. 1997. Maximizing the response of selection with a
447 predefined rate of inbreeding. *J. Anim Sci.* 75: 934-940.
- 448 Meuwissen, T. H., and J. A. van Arendonk. 1992. Potential improvements
449 in rate of genetic gain from marker-assisted selection in dairy cattle
450 breeding schemes. *J Dairy Sci* 75: 1651-1659.
- 451 Meuwissen, T. H. E., and M. E. Goddard. 1996. The use of marker
452 haplotypes in animal breeding schemes. *Genet. Sel. Evol.* 28: 161-176.
- 453 Pong-Wong, R., and J. A. Woolliams. 1998. Response to mass selection
454 when an identified major gene is segregating. *Genet. Sel. Evol.* 30: 313-
455 337.
- 456 Pong-Wong, R., and J. A. Woolliams. 2007. Optimisation of contribution
457 of candidate parents to maximise genetic gain and restricting inbreeding
458 using semidefinite programming. *Genet. Sel. Evol.* 39: 3-25.
- 459 Spelman, R. J., and J. A. M. van Arendonk. 1997. Effect of inaccurate

- 460 parameter estimates on genetic response to marker-assisted selection in
461 an outbred population. *J. Dairy Sci.* 80: 3399-3410.
- 462 Verrier, E. 2001. Marker assisted selection for the improvement of two
463 antagonistic traits under mixed inheritance. *Genet. Sel. Evol.* 33: 17-38.
- 464 Villanueva, B., S. Avendaño, and J. A. Woolliams. 2006. Prediction of
465 genetic gain from quadratic optimisation with constrained rates
466 of inbreeding. *Genet. Sel. Evol.* 38: 127-146.
- 467

TABLES AND FIGURES

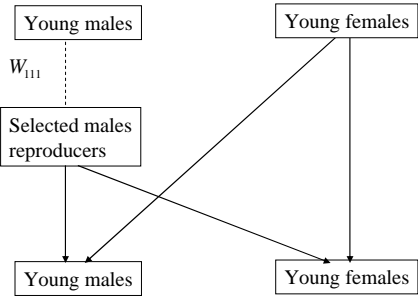


Figure 1. selection scheme with 3 categories (rectangles), one selection step (dashed line) and random mating (arrows).

479 **Table 1.** Optimised parameter values and objective function obtained with
480 the grid search and the genetic algorithm with a maximum loss of progress
481 of 10%.

Resolution	Generation number	θ_1	θ_2	Objective function
Grid search	5	-5.8	-8,8	0.43
	10	-4	-11,8	0.76
	15	-7	-12	0.94
Genetic algorithm	5	-5.23	-9.15	0.44
	10	-4.78	-10.8	0.77
	15	-6.98	-11.99	0.96

482

Table 2. Results obtained with the optimisation of the deterministic and the stochastic models with the optimised parameters θ_1, θ_2 and the objective function for different configurations: number of generation cycles, percentage of tolerated loss of genetic progress.

Model	Generation nb.	Accepted loss ΔG	θ_1	θ_2	Objective function
Deterministic	5	100%	-47.19	-90.11	0.83
		10%	-3.70	-8.20	0.44
	10	100%	-34.02	-99.89	0.99
		10%	-4.04	-9.94	0.78
	15	100%	-32.92	-97.70	1.00
		10%	-6.54	-11.42	0.99
	20	100%	-47.70	-57.30	1.00
		10%	-10.81	-16.26	1.00
Stochastic	5	100%	-40.67	-92.86	0.77
		10%	-5.23	-9.15	0.44
	10	100%	-37.33	-77.91	0.99
		10%	-4.78	-10.80	0.77
	15	100%	-37.74	-77.91	1.00
		10%	-6.98	-11.99	0.96
	20	100%	-60.71	-95.19	1.00
		10%	-9.37	-16.6	0.99

487

Table 3. Results obtained with different values of coefficients in the stochastic model: coefficients with the optimisation of the stochastic model and coefficients from the optimisation of the deterministic model

Generation nb.	Coefficients from the optimisation model	Objective function	Favourable genotype frequencies	% loss
Accepted loss =100%				
5	Deterministic	0.77	0.77	60%
	Stochastic	0.77	0.77	60%
10	Deterministic	0.99	0.99	33%
	Stochastic	0.99	0.99	33%
15	Deterministic	1.00	1.00	22%
	Stochastic	1.00	1.00	22%
20	Deterministic	1.00	1.00	16%
	Stochastic	1.00	1.00	16%
Accepted loss =10%				
5	Deterministic	0.32	0.2	9.3%
	Stochastic	0.44	0.44	10%
10	Deterministic	0.66	0.59	9%
	Stochastic	0.77	0.77	10%
15	Deterministic	0.90	0.87	9%
	Stochastic	0.96	0.96	10%
20	Deterministic	0.97	1	10.5%
	Stochastic	0.99	0.99	10%

Table 4. Objective function values obtained with the different indexes for 5, 10, 15 and 20 generations, 5 %, 10% and 100% of loss of genetic progress.

Generation number	Accepted loss ΔG	Objective function		
		1 st index	2 nd index	3 rd index
5	100%	0.768	0.768	0.768
	10%	0.440	0.442	0.441
	5%	0.368	0.370	0.368
10	100%	0.992	0.992	0.992
	10%	0.770	0.772	0.768
	5%	0.607	0.61	0.611
15	100%	1	1.00	1.00
	10%	0.950	0.965	0.963
	5%	0.802	0.807	0.803
20	100%	1.00	1.00	1.00
	10%	1.00	1.00	1.00
	5%	0.932	0.94	0.941

Quatrième partie

Modèle épidémio-génétique

Chapitre 6

Combinaison Modèle génétique-Modèle épidémiologique

6.1 Résumé

Afin de quantifier si l'optimisation de la sélection simultanée sur le gène PrP et sur un caractère de production peut avoir un réel impact sur l'évolution de la maladie, le modèle déterministe développé dans le chapitre 4 a été couplé avec un modèle dynamique épidémiologique. La combinaison de ces deux modèles permet d'étudier l'impact de différentes stratégies de sélection sur l'évolution de la tremblante. Les données utilisées pour effectuer nos simulations sont des données « réelles » issues de 3 races françaises : Lacaune, Manech Tête Rousse (MTR) et Manech Tête Noire (MTN)

Pour étudier l'impact de la sélection sur la propagation de la tremblante intra-troupeau dans les trois races (Lacaune, MTR et MTN), nous avons conçu plusieurs scénarios de simulation. Ces scénarios combinent trois stratégies de sélection (pas de sélection sur le gène PrP, sélection sur le gène PrP avec une forte contrainte sur la perte du progrès génétique du caractère de production (1%) et une sélection sur le gène PrP sans contrainte sur la perte du progrès génétique du caractère de production), à quatre scénarios d'introductions de la tremblante (nombre de brebis infectées au début de la simulation (1 ou 5) et nombre d'introduction de brebis infectés par an (0 ou 1)). Ces scénarios sont simulés sur 15 années.

Le modèle épidémiologique a pour objectif d'étudier la propagation de la tremblante classique au sein de différents troupeaux ovins. Ce modèle est basé sur un modèle SI-réalistes où la population est divisée en individus susceptibles de contracter la maladie (S), et en individus infectieux (I). Le modèle se concentre sur la population de brebis de remplacement. La contribution des béliers et des agneaux à la propagation de la maladie est négligée. En effet, le nombre de béliers au sein des troupeaux est faible et les agneaux qui ne sont pas gardés pour la reproduction sont abattus avant l'âge de huit mois. Ce modèle épidémiologique prend en compte la taille du troupeau, sa structure en âge et en génotype, le taux de renouvellement/réforme, le taux d'équarrissage et les politiques d'achats/ventes.

La combinaison du modèle déterministe génétique et du modèle épidémiologique se fait au niveau des accouplements dans le modèle épidémiologique. En effet, les fréquences génotypiques obtenues lors de l'optimisation de la sélection sur le gène PrP à chaque génération sont utilisées lors de la simulation des accouplements dans le modèle épidémiologique. Ces accouplements se font soit de manière déterministe soit de manière

stochastique.

Les résultats obtenus avec les accouplements considérés de manière déterministe ou stochastique sont en moyenne similaires. Nous avons décidé de nous attarder sur les résultats obtenus avec la prise en compte de la stochasticité.

Les résultats obtenus avec les différents scénarios montrent que l'incidence cumulée à la fin du processus est d'autant plus faible que la sélection sur le gène PrP est importante. Ce constat est le même quelque soit le troupeau considéré. Cependant il est plus marqué dans le troupeau où la fréquence initiale de l'allèle de résistance est faible. Ce constat est également visible au fur et à mesure des générations.

Cette différence se constate également au niveau des écart types de l'incidence de la maladie. Ces variables décroissent sous l'effet de la sélection contrairement aux fréquences génotypiques des animaux dit résistants qui augmentent. Ces résultats montrent que l'optimisation de la sélection sur le gène PrP pour la résistance à la tremblante a un effet très important sur l'incidence de la maladie et sur l'évolution de la structure de la population.

Les différents scénarios d'introduction de la tremblante peuvent être classés selon un ordre de « gravité » croissant : (1) introduction initiale d'un infecté sans introduction au cours du processus, (2) introduction initiale de cinq infectés sans introduction au cours du processus, (3) introduction initiale d'un infecté avec une introduction par an lors du processus et (4) enfin introduction initiale de cinq infectés avec une introduction par an lors du processus. Le cas où on considère initialement un infecté sans ajout de tremblante durant le processus se détache des autres scénarios. L'incidence cumulée est très nettement inférieure aux autres cas. Cette différence augmente sous l'effet de la sélection sur le gène PrP. En revanche pour les fréquences génotypiques, la différence est moins importante et quasiment inexistante au début du processus.

Pour vérifier la robustesse de nos résultats à l'incertitude sur les paramètres épidémiologiques, nous avons effectué une analyse de sensibilité. Le plan d'expérience considéré n'est pas complet mais satisfaisant puisque que le coefficient de détermination R^2 dans chaque analyse est supérieur à 99%. Cette analyse a permis de définir les paramètres ayant le plus d'impact (coefficient de transmission de la maladie, caractéristiques du troupeau, sélection, introduction de la tremblante). Les résultats obtenus permettent de mettre en évidence que l'incidence de la tremblante au sein d'un troupeau fluctue en fonction de leurs caractéristiques (fréquences génotypiques) ainsi que de la stratégie de sélection. Ces résultats sont venus conforter ceux obtenus précédemment.

Le troupeau « Lacaune » est un troupeau assez particulier (fréquence de l'allèle de résistance élevée dès le départ). Ces caractéristiques font que l'incidence cumulée se stabilise quelque soit l'introduction de la tremblante faite dans le troupeau. L'éradication de la maladie dans ce genre de troupeau s'avère plus facile que dans les autres troupeaux pour lesquels une sélection sur le gène PrP plus forte est nécessaire afin d'éradiquer la tremblante.

Les stratégies de sélection ont un fort impact sur l'épidémie de la maladie et sont donc un outil efficace pour le contrôle de la tremblante. Toutefois, leur efficacité en termes d'éradication de la maladie dépend fortement de la mise en place des scénarios épidémiologiques.

6.2 Article

Cet article est en cours de rédaction.

Combining dynamic epidemiological and genetic models to assess the impact of genetic selection strategies on the spread of classical scrapie within a sheep flock

Anne Costard, Suzanne Touzeau, Jean-Michel Elsen

Version 09/09/2010

1 Introduction

Scrapie is a fatal infectious neurodegenerative disease in sheep. Genetic resistance to clinical scrapie is associated with polymorphisms at three sites on the PrP gene, i.e. at codons 136(A/V), 154(R/H) and 171(Q/R/H) where each letter represents the amino acids at the indicate codon. The VRQ allele is associated with a high susceptibility to scrapie [10] while the ARR allele is associated with a resistance to scrapie and bovine spongiform encephalopathy (BSE) ([7]). It is difficult to distinguish between BSE and natural scrapie in sheep on the basis of clinical signs, so many countries have established eradication breeding program to eliminate scrapie in sheep breeds. These programs aim at increasing the ARR allele frequency in the populations. Data from flocks entering in such breeding programs were analysed. The analyses suggested that this introduction was an effective way to reduce the incidence of clinical scrapie confirmed by mathematical model ([15],[20]). Different mathematical epidemiologic modellings (deterministic or stochastic) were created to understand the assessing of susceptibility to scrapie, to characterize different transmission scenarii or to assess the effect of scrapie control measures ([15], [11], [14], [16], [2], [1], [9]).

However selection on PrP genotypes could have negative consequences on genetic progress for the other economical important traits. This may be the consequence of the loss of selection pressure, the emergence of a linkage disequilibrium between the PrP gene and the polygenic background, the increase of inbreeding and / or the loss of genetic variability. Some studies assessed those possible impacts of selection on PrP genotype([12], [3], [6], [13], [21]). Selection in favor of the ARR allele is expected to reduce the rate of genetic progress in selected trait in populations where the frequency of this allele is low. To help the control of this drawback, Costard et al ([5]) created a genetic deterministic and dynamic model to maximize scrapie resistance with a constraint on loss of genetic progress on an economical important trait such as milk production. It has been supposed that the disease does not become epidemic if the number of susceptible animals is low. Under this hypothesis, classical scrapie could be eradicated without the need to fix the ARR allele in the population ([15]).

The objective of this study was to evaluate the impact of selective breeding strategies on the spread of classical scrapie within French sheep flocks. We combined a dynamic epidemiological model with optimal outputs of a dynamic deterministic genetic model.

2 Materials and methods

2.1 Epidemiological model

The aim of this model is to represent the spread of classical scrapie within a sheep flock. The model focuses on the ewe population replacement. This is justified by the observation that few

rams are generally present in a flock and that lambs that are not kept for breeding purposes are culled before the age of eight months. The contribution of rams and lambs to the disease spread can therefore be neglected.

The realistic SI-like epidemiological model is based on [18] and [19], where S represents the susceptible and I the infected incubating individuals. The population representation is simplified, as it incorporates age classes instead of a continuous age structure. The main features are similar: seasonality in breeding and transmission, genetic and age-dependent susceptibilities, long and variable infectious incubation periods. The epidemiological and demographic processes are refined: two lambing seasons for ewes and ewe-lambs, selective breeding, stochastic mating, seasonal reform, variable exposure to the disease. The elements S,I,M,B represent the number of animals Infected, Susceptible, Dead or Born in a given period of time and within a genotype and an age class. They are represented in Figure 1 and described below. The complete set of equations appears in Appendix A.

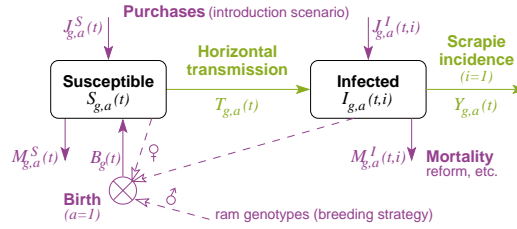


Figure 1: Epidemiological model. Flock at time t of susceptible (S) and infected incubating (I) renewal ewes, structured by genotype ($g \in \{1, \dots, N_g\}$), age class ($a \in \{1, \dots, N_a\}$), and infection load ($i \in [0, 1]$). Demographic processes (purple arrows) include birth (B), represented by an inflow of susceptible individuals in the first age class ($a = 1$), non scrapie mortality (M) and purchases (J). Epidemiological processes (green arrows) include horizontal transmission (T) and scrapie incidence (Y), which occurs at the end of the incubation period ($i = 1$). Birth depends on the breeding strategy that determines the ram genotypes; purchases depend on the introduction scenario.

Birth There are two lambing seasons: the main one for ewes of age two and more that gives birth to replacement ewes ($s_b(t) = 1$ if time $t \in$ lambing season); another one for the ewe-lambs whose lambs are not integrated in the study population (*idem* with $s_l(t)$).

To determine the breeding matrix ($G_{gg'}$) which gives the proportion of lambs with genotype g born from ewes with genotype g' , one needs to know the genotypes of the rams used for mating. It depends on the breeding strategy chosen (cf. Section 2.3.2). When no selective breeding is implemented, the ram genotype frequencies are similar to those of the flock. Otherwise, they are obtained by the genetic model (cf. Section 2.2). As few rams are used for mating, their genotypes are drawn from a multinomial law, the probabilities being given by the genotype frequencies.

Breeding aims at maintaining the total population size P at a given level, so the birth rate b is adjusted so as to reach a target population size at the end of the lambing season. The birth function is hence given by

$$B_g(t) = s_b(t)b(t, P) \sum_{g'} G_{gg'}(t) \sum_{a' \geq 2 \text{ yr}} P_{g',a'}(t), \quad (1)$$

where $P_{g',a'}(t) = S_{g',a'}(t) + \int_0^1 I_{g',a'}(t, i) di$ represents the population in age class a' , with genotype g' , at time t .

Horizontal transmission Horizontal transmission concerns all age classes and includes perinatal transmission. Vertical transmission is neglected, as contamination in lambs often occurs shortly after birth rather than *in utero* [4].

The PrP genotype is a major susceptibility gene for scrapie, so relative genetic susceptibilities ($\sigma_g \in [0, 1]$) are introduced in the model. Young animals are considered to be more at risk. The relative age susceptibilities ($\sigma_a \in [0, 1]$) therefore follow a decreasing step function.

Transmission is assumed to take place mainly during lambing seasons (scaling factor k) [19]. It is hence seasonal, even if some residual contamination may occur all year round (scaling factor $k_r \ll k$). Moreover, exposure varies among animals according to their age class: during the main lambing season ($s_b(t) = 1$), the ewes and their lambs are at greater risk ($e_b = 1$ for $a > 2$ years and $a < 6$ months) than the ewe-lambs ($e_b = \underline{e} < 1$), which are present in the flock but not in the lambing pens.

The residual force of infection is assumed to be proportional to the total number of infected sheep ($I(t)$). Similarly, the force of infection during each lambing period is proportional to the number of infected ewes that are lambing during this period ($I_b(t)$ or $I_l(t)$). The horizontal transmission outflow is hence given by

$$T_{g,a}(t) = \sigma_g \sigma_a S_{g,a}(t) [k_r I(t) + k s_b(t) e_b(a) I_b(t) + k s_l(t) e_l(a) I_l(t)],$$

where $S_{g,a}(t)$ represents the susceptible population in age class a , with genotype g , at time t .

Scrapie incubation and incidence Once an individual is contaminated, it enters a long and undetectable incubation period (*ante-mortem* tests have a moderate efficacy). In the model, it acquires a first infection load $i \in [0, 1]$ that grows exponentially during the incubation period until it reaches its maximum value $i = 1$. This value corresponds to the onset of clinical signs, very shortly followed by the removal of the individual from the flock.

The exponential growth rate (c_g) of the infection load depends on the sheep PrP genotype g , so the length of the incubation period varies among genotypes. Moreover, to allow some more variability in the incubation period, the first infection load of newly infected individuals follow a Beta distribution $\beta(i)$. Let i_0 be the first infection load. We have $\frac{di}{dt} = c_g i$. So the length of the incubation period is $\tau_g(i_0) = -\log(i_0)/c_g$.

The cumulated incidence at time t corresponds to the number of scrapie cases between time 0 and time t . It is hence

$$\mathcal{Y}_{g,a}(t) = \int_0^t Y_{g,a}(x) dx = \int_0^t c_g I_{g,a}(x, 1) dx, \quad (2)$$

where $I_{g,a}(x, 1)$ represents the density of infected sheep in age class a , with genotype g and infection load $i = 1$, at time x .

Reform and non-scrapie mortality Reform is seasonal (it usually occurs after mating) and age-dependent. As infected individuals are silently incubating in the flock, it does not depend on the scrapie status. Moreover, we assume that the reform rate is genotype-independent ($\mu_a(t)$), either because the flock is not genotyped, or because selection is only applied on breeding and not culling.

An extra constant death rate (ν) is introduced to represent mortality due to injuries or diseases other than scrapie. The mortality outflow for susceptible ewes is hence

$$M_{g,a}^S(t) = (\mu_a(t) + \nu) S_{g,a}(t) \quad (3)$$

and is similarly defined for infected ewes.

Purchases The flock is open, meaning that animals are bought in and sold out of the flock. The purchase of animals represents a potential scrapie risk and is explicit in the model. Animals that are sold represent a potential risk for other flocks. As this model is flock-centred, they are not distinguished from animals that are reformed.

The time at which ewes are bought in the flock, their age and their genotype depend on flock management. The ratio of infected individuals among the purchases is determined by the introduction scenario (cf. Section 2.3.2).

2.2 Genetic model

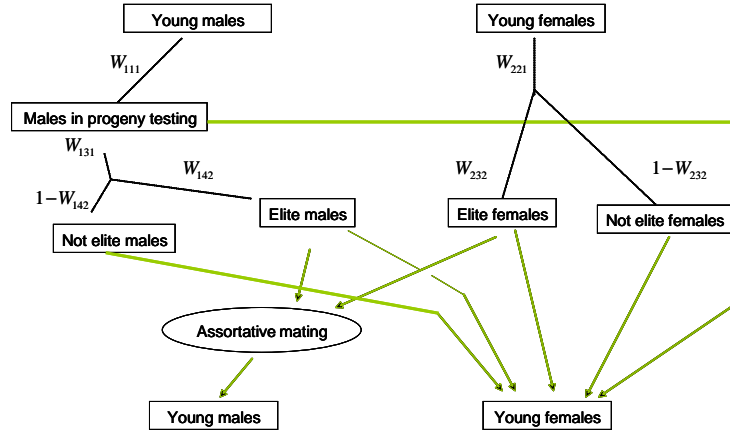


Figure 2: Genetic model. Population description with 7 categories (rectangles), 7 matings paths (green lines), selection steps (dark lines) and selection proportion W_{sac} (sex $s=1$: males, 2: females; age $a=1, \dots, a_s$; category $c=1$: not elite, 2: elite).

A deterministic mathematical approach, based on a realistic and fairly detailed selection model represented in Figure 2, was developed to globally optimise the selection scheme. It simultaneously optimize selection on two traits, a quantitative trait with underlying polygenic variation and a monogenic trait [5]. It maximizes the frequency of the desired genotype for the monogenic trait, while minimizing the loss of genetic progress on the polygenic trait. Breeding programs with overlapping generations, different selection strategies for males and females, and assortative mating were modeled. This modeling approach may easily be adapted to a variety of underlying genetic models and selection schemes. It was applied here to select on the PrP gene for scrapie resistance in a dairy sheep selection scheme.

Description of the population A population of infinite size with overlapping generations and different classes of animals defined by sex s ($s = 1$ for males and $s = 2$ for females), age a , and estimated breeding value (EBV) categories c (elite, not elite) was considered. According to

their mean EBV, animals from identical sex and age may be classified in different subsets, which determine their use as reproducers. Two alleles (favorable vs. unfavorable) were considered for the PrP gene, thus three genotypes exist ($g = 1$ favourable genotype, $g = 2$ intermediate genotype, and $g = 3$ unfavourable genotype). Males and females are aged from 0 to 6 years. Classes are characterised at each generation t by their genotype frequencies (f_{sactg}) and their EBV (μ_{sactg}).

Selection Females were selected on their own performances (with a selection rate $W_{sac} = W_{221}$) after their 1st lactation between 2 and 3 years of age. The selected females were distributed in 2 categories, elite (W_{232}) and non elite ($1 - W_{232}$). Among the young males born from matings between elite reproducers, a fraction W_{111} was mated to non elite females and their EBVs for the polygenic trait were estimated after a progeny test. A fraction W_{131} of these males was selected 2 years later. Among the selected males, a proportion was mated to the elite females to produce young male candidates for selection in the next generation. The replacement of females resulted from mating between (i) non elite females and elite males, non elite males, and males in progeny testing; (ii) elite females and elite males. PrP genotypes were known for females and males. Only males were selected on their PrP genotype and on their EBV value before and after progeny testing.

Optimisation The aim was to maximize scrapie resistance, minimizing the loss of genetic progress in milk production. The objective function chosen was the frequency of homozygous favourable females at the end of the process ($s = 2$, $g = 1$, $t = T$) with a penalty on the loss of genetic progress. The penalty prevents the progress from being lower than a given target value, which corresponds to a percentage π of the progress obtained without scrapie selection (ΔG_T). The objective function is written as follows:

$$F_{obj} = \frac{\sum_{a=1}^6 \sum_{c=1}^2 \frac{1}{a} f_{2acT1}}{\sum_{a=1}^6 \frac{1}{a}} - 100 * \frac{\pi \Delta G_T - \sum_{g=1}^3 f_{211Tg} \mu_{211Tg}}{\pi \Delta G_T} \quad (4)$$

The first term is the mean of favourable ($g = 1$) elite ($c = 1$) and non elite ($c = 2$) females, weighted by the $\frac{1}{a}$ coefficient which gives more importance to younger females. All females were considered in the objective function because they all participate in the renewal. The second term corresponds to the penalty on the loss of genetic progress. A genetic algorithm [8] was used to solve this optimization problem.

Ram PrP genotype frequencies obtained thanks to this optimization procedure were used as an input for the breeding matrix ($G_{gg'}$) of the epidemiological model described in Section 2.1.

2.3 Simulation scenarios

2.3.1 Model parameters

The model was applied to three French dairy flocks characterised by their sheep breed: Lacaune from the Roquefort region (LAC), Manech Tête Noire (MTN) and Manech Tête Rousse (MTR), both from the Pyrénées Atlantiques. These three breeds exhibit contrasted PrP genotype frequencies that are shown in Table 1.

Further data were available for each flock: flock size, age pyramid, number of rams, lambing seasons (for ewes and ewe-lambs), reform rate, and mortality rate. They were provided by the Institut de l'Élevage and INRA and reflect a typical Lacaune, MTN or MTR flock. They were used to estimate the demographic parameters of the model.

The epidemiological parameters are shown in Table 2. Reasonable values were chosen for the relative susceptibilities of the RS genotype and of older animals (50% compared to the

Flock	Genotype		
	$g = 1$: RR	$g = 2$: RS	$g = 3$: SS
LAC	0.43	0.48	0.09
MTN	0.19	0.53	0.28
MTR	0.05	0.37	0.58

Table 1: PrP genotype frequencies for the three flocks characterised by their sheep breed: La-caune (LAC), Manech Tête Noire (MTN) and Manech Tête Rousse (MTR). R denotes the scrapie resistant allele and S the susceptible allele.

<i>Transmission parameters</i>		
Relative genetic susceptibilities	$\sigma_{RR} = 0$	
	$\sigma_{RS} = 0.5$	
	$\sigma_{SS} = 1$	(scaled on SS genotype)
Relative age susceptibilities	$\sigma_{a \geq 2 \text{ yrs}} = \underline{\sigma} = 0.5$	
	$\sigma_{a < 2 \text{ yrs}} = 1$	(scaled on younger ewes)
Exposure coefficients (season s_b or s_l)	$e_b \text{ or } l = \underline{e} = 0.25$	for ewes not lambing
	$e_b \text{ or } l = 1$	(scaled on ewes lambing)
Scaling factors	$k = 8.4 \cdot 10^{-2} \text{ yr}^{-1} \cdot \text{indiv}^{-1}$	seasonal transmission
	$k_r = 1.7 \cdot 10^{-4} \text{ yr}^{-1} \cdot \text{indiv}^{-1}$	residual transmission
<i>Incubation parameters</i>		
First infection load	$i_m = 0.1$	mean of Beta distribution
	$i_v = 8.3 \cdot 10^{-4}$	variance —
	$\Leftrightarrow \beta_1 = 10.7, \beta_2 = 96.3$	parameters —
Infection load growth rate	$c_{RR} = 0 \text{ yr}^{-1}$	
	$c_{RS} = 1.15 \text{ yr}^{-1}$	
	$c_{SS} = 1.35 \text{ yr}^{-1}$	
\Leftrightarrow Incubation period length	$\tau_{RS} = 2 \text{ yrs}$	(computed with
	$\tau_{SS} = \underline{\tau} \tau_{RS} = 1.7 \text{ yrs}$	first infection load i_m)

Table 2: Epidemiological parameters.

most susceptible groups), as well as for the exposure coefficient (25% for the animals present in the flock but not in the lambing pens). The incubation parameters for the SS genotype were obtained from [19]. Assuming a mean first infection load, it corresponds to an incubation period of 1.7 years. The parameters for the RS genotype were similarly chosen, assuming a 2-year incubation period. The scaling coefficients were calibrated in the model so as to obtain a very low scrapie incidence due to residual transmission and “reasonable” values from the seasonal transmission during lambing periods. As these parameter values remain fairly uncertain, a sensitivity analysis was conducted to test the robustness of our results (cf. Section 2.4).

2.3.2 Scenarios

To study the impact of selective breeding on the scrapie spread in the three flocks selected above, we designed several simulation scenarios. They are based on the breeding strategy and the scrapie introduction scenario. Three breeding strategies were implemented:

NS no PrP selection, i.e. the ram PrP genotypes are similar to those of the ewes;

SC1 PrP selection on the ram genotypes, with a strong constraint on the loss of genetic progress in milk production ($\pi = 1\%$);

SNC PrP selection on the ram genotypes, without constraint on the loss of genetic progress in milk production.

These three strategies were combined with four introduction scenarios obtained from:

- the initial number of infected ewes in the flock, $I^0 = 1$ or 5 ewes;
- the yearly number of infected ewes bought in the flock, $I^1 = 0$ or 1 ewe per year.

We assume that these are young ewes, with genotype RS, that are just starting their incubation period when introduced in the flock.

We hence obtain 3 flocks \times 3 breeding strategies \times 4 introduction scenarios, i.e. 36 simulation scenarios. The ram genotypes being drawn from a multinomial law, the model is stochastic. We therefore performed 150 replications for each scenario, leading to 5400 simulations. We chose a 15-year time horizon, which more or less corresponds to two generations. We used the Scilab¹ software.

2.3.3 Outputs

We chose to look at two model outputs: (i) the dam and ram PrP genotype frequencies over time, which show the impact of selective breeding on the flock genetic structure; (ii) the total scrapie cumulated incidence over time, $\mathcal{Y}(t) = \sum_{g,a} \mathcal{Y}_{g,a}(t)$, which relates to the spread of the disease.

Note that both outputs can actually be observed in a flock. PrP genotypes can be determined from a blood sample. The incidence corresponds to scrapie cases, i.e. animals that show clinical signs and whose post-mortem tests are positive. The infected population, however, cannot be observed: it corresponds to animals incubating the disease and therefore cannot be distinguished from the susceptible population.

2.4 Sensitivity analysis

To check the robustness of our results to the uncertainty on the epidemiological parameters, we performed a sensitivity analysis. It aims at defining which parameters have the greatest impact on a given model output.

The output we chose is the total cumulated incidence at the end of the simulation period $\mathcal{Y}(T)$, with $T = 15$ years. It is an aggregated variable that relates to scrapie spread.

We included all epidemiological parameters had them vary over a $\pm 20\%$ range around their nominal values given in Table 2. We hence obtained eight parameters: σ_{RS} , $\underline{\sigma}$, \underline{e} , k , k_r , β_1 , τ_{RS} , and $\underline{\tau}$. We chose $\underline{\tau} \in [0, 1]$ instead of τ_{SS} so as to ensure that $\tau_{SS} \leq \tau_{RS}$ which is biologically consistent. With little loss of generality, the mean of the first infection load i_m can be fixed; the remaining incubation parameters still generate a wide range of incubation period distributions. We chose to vary β_1 , fixing $\beta_2 = 9\beta_1$ ($\Leftrightarrow i_m = 0.1$).

So as to determine the interactions between the epidemiological parameters and the simulation scenarios, we also included the scenario parameters defined in Section 2.4 in the analysis: $\text{sel} \in \{\text{NS}, \text{SC1}, \text{SNC}\}$, $I^0 \in \{1, 5\}$, and $I^1 \in \{0, 1\}$. The analysis was performed for each flock, but also with the flock as a parameter. It allows to include the demographic parameters, especially the PrP genotype frequencies of Table 2, as one meta-parameter in the analysis. In total, we hence have 11 or 12 parameters.

A factorial experimental design was used to explore the parameter space, with three levels for the epidemiological parameters (nominal value & $\pm 20\%$). A complete design would lead to

¹Scilab, the Free Platform for Numerical Computation: <http://www.scilab.org/>

$3 \times 2^2 \times 3^8 = 78732$ simulations per flock. So we used a fractional design based on the following statistical model:

$$([\text{flock+}]_{\text{sel}} + I^0 + I^1)^3 \cdot (\sigma_{RS} + \underline{\sigma} + \underline{\epsilon} + k + k_r + \beta_1 + \tau_{RS} + \underline{\tau})^2$$

It allows the estimation of main effects and interactions involving up to three scenario parameters, combined with up to two epidemiological parameters. The higher order interactions are assumed to be negligible. It leads to 2916 simulations/flock [8748 simulations when flock is a parameter]. The number of simulations remains fairly high, so we chose to perform the analyses on the deterministic version of the model, in which the PrP genotypes of the rams directly correspond to the genotype frequencies.

An ANOVA was then performed, to compare the influence of the 11[12] parameters on the output. Sensitivity indices and total sensitivities were calculated for each parameter. The sensitivity index SI_x is defined as the ratio between the sum of squares associated with the main effect of parameter x and the total sum of squares. The total sensitivity TS_x is given by the sum of the sensitivity indices corresponding to parameter x (main effect and interactions) [17].

The analyses were performed with R².

3 Results

The results with the determinist and stochastic models were on average similar. In this section, we discuss the results obtained with the stochastic model.

Results on total cumulated incidence with the worst introduction scenario ($I^0 = 5$ and $I^1 = 1$). After 15 years with the auto renewal, we obtained a cumulated incidence of 36.07, 154.28 and 209.41 for initial animal frequencies with RR genotypes of 0.43, 0.19 and 0.05 in the different flocks (Lacaune, MTN and MTR respectively). The cumulated incidence was multiplied by 4 between Lacaune and MTN flocks and by 6 between Lacaune and MTR flocks. At the end of process with the auto renewal, for each flock, the SS genotype animal frequencies were of 0.1, 0.16 and 0.43 respectively. The importance of initial conditions in flocks on disease incidence was therefore demonstrated by these results. The largest cumulated incidence (209.41) was obtained in the MTR flock with the auto renewal. With a PrP selection with a strong constraint on loss of genetic progress (loss 1% of genetic progress in milk production), we noticed a decrease of cumulated incidence (154.01). This decrease was even more important (68.01) with a PrP selection and no constraint in the loss of genetic progress. With the effect of the PrP selection, the total cumulated incidence was thus divided by 3. These results were also observed in MTN and Lacaune flocks where the total cumulated incidence was divided approximately by 2.5 and by 1.9 respectively.

Results on genotype frequencies with the worse introduction scenario ($I^0 = 5$ and $I^1 = 1$). The PrP selection effect was also noticed on the different genotype frequencies. In the MTR flock without PrP selection and with the worst scenario, the SS genotype frequency was of 0.43. With a PrP selection and a low constraint on the loss of genetic progress, the SS genotype frequency decreased (0.20). This decrease was even more important when there was no constraint on the loss of genetic progress (SS genotype frequency of 0.01). These results were confirmed in MTN and Lacaune flocks where the SS genotype frequencies were of 0.16 and 0.10 respectively, then decreased towards 0.09 and 0.04 and finished stabilized at 0.006 and 0.003. The standard deviations of the total cumulated incidence decreased under the PrP selection influence. Indeed, in the MTR flock, the standard deviations decreased from 21.27 to 3.55 at

²The R Project for Statistical Computing: <http://www.r-project.org/>

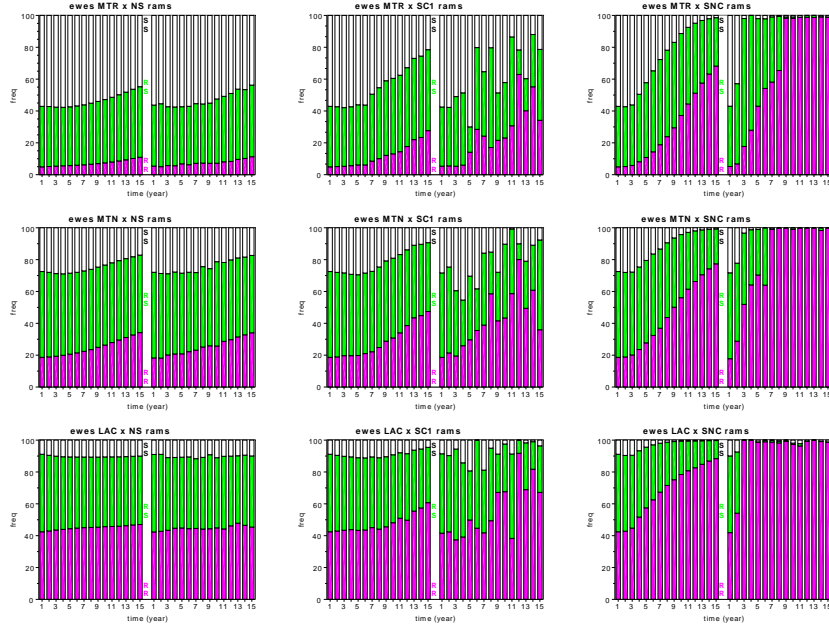


Figure 3: PrP genotype frequencies for the three flocks (MTR, MTN, LAC; in lines) and the three breeding strategies (NS, SC1, SNC; in columns): RR (purple), RS (green), and SS (white) genotypes; ewes on the left, rams on the right of each subfigure. Introduction scenario: $I^0 = 5$ initial infected ewes + $I^0 = 1$ infected ewe per year.

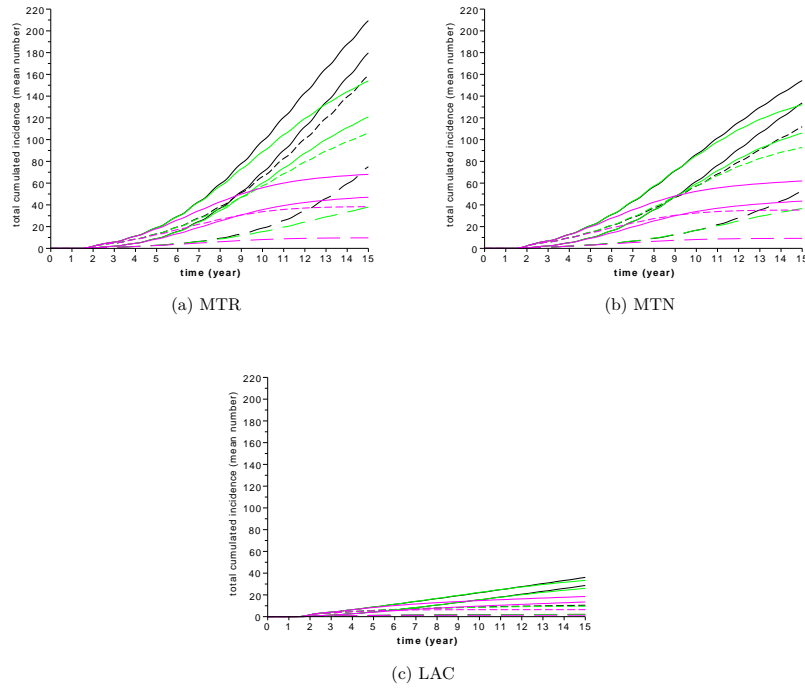


Figure 4: For each flock (1 flock per subfigure): total cumulated incidence over time, computed for the 12 scenarios (mean number computed over the 150 replications per scenario). The scenarios are defined by the 3 breeding strategies: no selection (NS, black), selection for scrapie resistance with less than 1% loss on genetic progress (SC1, green), selection without constraint (SCN, purple) \times the 4 introduction scenarios: $I^0 = 1$ (thin line) or 5 (thick line) initial infected ewes + $I^1 = 0$ (dashed line) or 1 (plain line) infected ewe per year.

the end of the process (15 years). In the MTN flock, they fluctuated between 20.96 and 3.62. In the Lacaune flock, they fluctuated between 7.64 and 0.84. This difference was also confirmed with the genotype frequencies but in a less obvious manner.

Classification of different introduction scenarii The different scenarii of scrapie introduction could be classified according to an order of "gravity": $I^0 = 1$ and $I^1 = 0$, $I^0 = 5$ and $I^1 = 0$, $I^0 = 1$ and $I^1 = 1$ and $I^0 = 5$ and $I^1 = 1$. The " $I^0 = 1$ and $I^1 = 0$ " scenario differed from the others scenarii. Indeed, the total cumulated incidence was much lower than the other scenarii. For example, in the MTR flock with auto-renewal, the total cumulated incidence was of 75.08 with the scenario with $I^0 = 1$ and $I^1 = 0$, 159.03 with the scenario $I^0 = 5$ and $I^1 = 0$, 179.75 with the scenario $I^0 = 1$ and $I^1 = 1$ and 209.41 with the scenario $I^0 = 5$ and $I^1 = 1$. These results were similar for the MTN and Lacaune flocks. This difference increased under the PrP selection. Without the PrP selection, the total cumulated incidence was multiplied by 3 between the scenarii $I^0 = 1$ and $I^1 = 0$ and $I^0 = 5$ and $I^1 = 1$. With a low PrP selection, the total cumulated incidence was multiplied by 4. With no constraint on the loss of genetic progress, it was multiplied by 7. These contrasts between scenarii $I^0 = 1$ and $I^1 = 0$ and $I^0 = 5$ and $I^1 = 1$ were the same in MTN flock. For the Lacaune flock, these ratios were more important (17, 16.5 and 14) and the differences were less important for the genotype frequencies. The total cumulated incidence at the end of the processus, in all flocks, was stabilised with PrP selection and no constraint on loss of genetic progress. In the other scenarii, the total cumulated incidence continued to grow (Figure 4) excepted in Lacaune flock. In Lacaune flock, the total cumulated incidence stabilized whatever scrapie introduction was made (Figure 4c).

Sensitivity analysis At first, the sensitivity analysis was realized with all flocks. The factors influencing most the variation of total cumulated incidence were in this order: k , $flock$, sel , I^1 , τ_{rs} , I^0 , σ_{rs} . The results obtained before showed the importance of flocks. Indeed, the difference between flocks were very important, so the sensitivity analysis was realized by flocks to see if factors influencing the sensitivity analysis were modified. For each analysis, the coefficient of determination R^2 was upper in 99%. Consequently, the considered experiment plan was sufficient, the complete plan was not necessary. The factors influencing the variation of total cumulated incidence, at the end of process, were different according to flocks. For the Lacaune flock, the most important factors were I^1 , sel , k , σ_{rs} and I^0 . These factors explained 95% of variation of total cumulated incidence. For the MTR flock, k , sel , τ_{rs} , I^1 , I^0 were the most important factors explaining 66% of variation of total cumulated incidence and for the MTN flock, k , sel , I^1 , τ_{rs} , σ_{rs} explaining 70% of variation of total cumulated incidence. k , sel , I^1 were the 3 factors common to the 3 flocks. For a fixed transmission coefficient, these results coincided with those of the previous section 3. These results put in evidence that the total cumulated incidence of scrapie within a flock fluctuated according to the initial conditions as well as to the mating strategy.

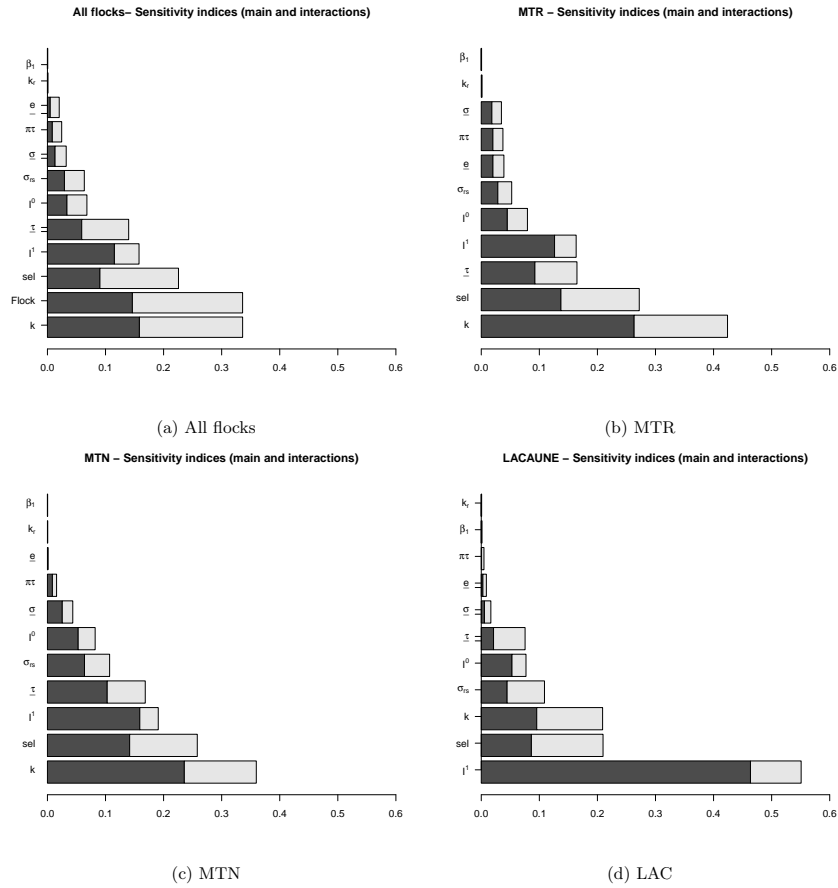


Figure 5: Sensitivity analyses on the total cumulated incidence: results of the ANOVA (a) for all flocks, flock being a parameter; and (b)-(d) for each flock. Total sensitivities (TS_x) are represented for the 4 (a) or 3 (b)-(d) scenario parameters and the 8 epidemiological parameters. They are ranked in descending order and split in main effect (dark grey – SI_x sensitivity index) and interactions (light grey).

4 Discussion

A dynamic epidemiological model was combined with optimal outputs of a dynamic deterministic genetic model to evaluate the impact of selective breeding strategies on the spread of classical scrapie within French sheep flocks. In this study, a low level of stochasticity in breeding and non-existent for the contamination was used. Therefore, the infection is overestimated (noextinction in finite time), which is quite realistic for scrapie. The data used in this study came from three main French dairy flocks. These flocks had contrasting characteristics, including their genotypic structure. Our model takes into account many details: selection rate, genotypic frequencies, selection accuracy, size, and age and genotypic structure of flocks, culling rate. Despite epidemiological parameters uncertain reinforced by a sensitivity analysis, our results appear robust. The mean and standard deviation of the total cumulated incidence and SS genotype frequencies decreased due to the PrP selection. Therefore, the optimisation of the PrP selection to scrapie resistance has an important impact on disease incidence and on the evolution of population structure. Breeding strategies have a strong impact on the disease outbreak and are therefore an efficient tool to control scrapie. However, their efficacy in terms of disease eradication strongly depends on the scrapie introduction scenarios and initial genotypic structure of flocks. Therefore, the disease eradication in flocks similar to the Lacaune flock seemed to be easier than in the other flocks, for which selection was needed to eradicate scrapie. However, in four generations, a population resistant to the introduction of scrapie can be obtained. By acting only on the rams, the results are really satisfactory. These results can be improved using assortative mating. To evaluate the evolution of the disease and the population structure, we have chosen to analyze the Total Cumulated incidence and genotype frequencies. The age structure, which has been taken into account in these models, has a significant impact on the epidemiology the total cumulated incidence corresponds to a synthetic indicator. The incidence corresponds to observations, the newly infected are off by one year compared to the incidence. The incidence is one of the most important criteria to evaluate the frequency and speed of onset of disease. The simulations are over 15 years as Man et al (2009). This disease has a long incubation period (two years) and the time to completely renew the generations is long enough (four years). In reality, these scenarios do not exist, stringent sanitary measures have been put in place. However, in this study, the objective is to observe the evolution of the disease without sanitary measures with only selection strategies, which explains high incidence. These results hence emphasise the contribution of our original approach, which combines genetic and epidemiological models in a realistic framework.

References

- [1] Hagenaars T. J. Donnelly C. A. and Ferguson N. M. Epidemiological analysis of data for scrapie in Great Britain. *Epidemiol. Infect.*, 134:359–367, 2006.
- [2] Hagenaars T. J. Ferguson N. M. Donnelly C. A. and ANDERSON R. M. Persistence patterns of scrapie in a sheep flock. *Epidemiology and Infection*, 127:157–167, 2001.
- [3] Leopoldo Alfonso, Analia Parada, Andres Legarra, Eva Ugarte, and Ana Arana. The effects of selective breeding against scrapie susceptibility on the genetic variability of the Lacaune black-faced sheep breed. *Genetics Selection Evolution*, 38(5):495 – 511, 2006.
- [4] Olivier Andréoletti, Caroline Lacroux, Armelle Chabert, Laurent Monnerneau, Guillaume Tabouret, Frédéric Lantier, Patricia Berthon, Francis Eychenne, Sylvie Lafond-Benestad, Jean-Michel Elsen, and François Schelcher. PrPSc accumulation in placentas of ewes exposed to natural scrapie: influence of foetal PrP genotype and effect on ewe-to-lamb transmission. *J. Gen. Virol.*, 83(10):2607–2616, 2002.

- [5] A. D. Costard, Z. G. Vitezica, C. R. Moreno, and J.-M. Elsen. A dynamic deterministic model to optimize a multiple-trait selection scheme 10.2527/jas.2008-0898. *J. Anim Sci.*, 87(3):885–894, 2009.
- [6] C. Drogemuller, F. De Vries, H. Hamann, T. Leeb, and O. Distl. Breeding german sheep for resistance to scrapie. *Vet Rec.*, 154(9):257–260, 2004.
- [7] J. D. Foster, D. Parnham, A. Chong, W. Goldmann, and N. Hunter. Clinical signs, histopathology and genetics of experimental transmission of bse and natural scrapie to sheep and goats. *Vet Rec.*, 148(6):165–171, 2001.
- [8] D.E. Goldberg. *Genetics algorithms in search, optimization and machine learning*. Addison-Wesley Publishing Company, Inc., Boston, 1989.
- [9] S. Gubbins. Transmission of scrapie between sheep flocks in great britain. *Proceedings of the 10th Symposium of the International Society for Veterinary Epidemiology and Economics, Vina del Mar, Chile.*, ISVEE 10:725, 2003.
- [10] N. Hunter, J.D. Foster, W. Goldmann, M.J. Stear, J. Hope, and C. Bostock. Natural scrapie in a closed flock of cheviot sheep occurs only in specific prp genotypes. *Archives of Virology*, 141:809–824, 1996.
- [11] Matthews L. Woolhouse M. E. J. and Hunter N. The basic reproduction number for scrapie. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1423):1085–1090, 1999.
- [12] Wing-Young Man, Ronald Lewis, Kay Boulton, and Beatriz Villanueva. Predicting the consequences of selecting on prp genotypes on prp frequencies, performance and inbreeding in commercial meat sheep populations. *Genetics Selection Evolution*, 39(6):711 – 729, 2007.
- [13] A. Molina, M. Juárez, and A. Rodero. Merino sheep breed’s genetic resistance to scrapie: Genetic structure and comparison of five eradication strategies. *Preventive Veterinary Medicine*, 75(3-4):239–250, 2006. TY - JOUR.
- [14] Matthews L. Coen P. G. Foster J. D. Hunter N. and Woolhouse M. E. J. Population dynamics of a scrapie outbreak. *Archives of Virology*, 146, Number 6 / juin 2001:1173–1186, 2001.
- [15] Woolhouse M. E. J. Stringer S. M. Matthews L. Hunter N. and Anderson R. M. Epidemiology and control of scrapie within a sheep flock. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1402):1205–1210, 1998.
- [16] Hagenaars T.J. Donnelly CA. Ferguson NM. and Anderson RM. The transmission dynamics of the aetiological agent of scrapie in a sheep flock. *Math Biosci.*, 168 (2):117–135, 2000.
- [17] Andrea Saltelli, K. Chan, and E. M. Scott, editors. *Sensitivity Analysis*. Wiley series in Probability and Statistics. John Wiley & Sons, 2000.
- [18] S. M. Stringer, N. Hunter, and M. E. J. Woolhouse. A mathematical model of the dynamics of scrapie in a sheep flock. *Math. Biosci.*, 153(2):79–98, 1998.
- [19] S. Touzeau, M. E. Chase-Topping, L. Matthews, D. Lajous, F. Eychenne, N. Hunter, J. D. Foster, G. Simm, J.-M. Elsen, and M. E. J. Woolhouse. Modelling the spread of scrapie in a sheep flock: evidence for increased transmission during lambing seasons. *Arch. Virol.*, 151(4):735–751, April 2006.

- [20] C. Webb, C.R Sauter-Louis. Investigatios inti thr contact structure of the british sheep pop-
ulation. *Proceedings of the Society for Veterinary Epidemiology and Preventive Medecine*,
Cambridge, pages 10–20, 2002.
- [21] Moll L. Windig J. J., Eding H. and Kaal L. Effects on inbreeding of different strategies aimed
at eliminating scrapie sensitivity alleles in rare sheep breeds in the netherlands. *Animal*
Science, 79 Part 1:11–20, 2004.

A Epidemiological model

$$\begin{aligned}\frac{dS_{g,1}(t)}{dt} &= s_b(t)b(t, P) \sum_{g'=1}^{N_g} G_{gg'}(t) \sum_{a' \geq 2yr} \left[S_{g',a'}(t) + \int_0^1 I_{g',a'}(t, i') di' \right] \\ &\quad - (\alpha_1 + \mu_1(t) + \nu) S_{g,1}(t) - \sigma_{g,1} S_{g,1}(t) \phi_1(t, I) \\ \frac{dS_{g,a}(t)}{dt} &= J_{g,a}^S(t) + \alpha_{a-1} S_{g,a-1}(t) - (\alpha_a + \mu_a(t) + \nu) S_{g,a}(t) \\ &\quad - \sigma_{g,a} S_{g,a}(t) \phi_a(t, I)\end{aligned}\tag{5}$$

$$\begin{aligned}\frac{dS_{g,N_a}(t)}{dt} &= \alpha_{N_a-1} S_{g,N_a-1}(t) - (\mu_{N_a}(t) + \nu) S_{g,N_a}(t) \\ &\quad - \sigma_{g,N_a} S_{g,N_a}(t) \phi_{N_a}(t, I) \\ \left(\frac{\partial}{\partial t} + \frac{\partial c_g i}{\partial i} \right) I_{g,1}(t, i) &= - (\alpha_1 + \mu_1(t) + \nu) I_{g,1}(t, i) + \beta(i) \sigma_{g,1} S_{g,1}(t) \phi_1(t, I) \\ \left(\frac{\partial}{\partial t} + \frac{\partial c_g i}{\partial i} \right) I_{g,a}(t, i) &= J_{g,a}^I(t, i) + \alpha_{a-1} I_{g,a-1}(t, i) \\ &\quad - (\alpha_a + \mu_a(t) + \nu) I_{g,a}(t, i) + \beta(i) \sigma_{g,a} S_{g,a}(t) \phi_a(t, I) \\ \left(\frac{\partial}{\partial t} + \frac{\partial c_g i}{\partial i} \right) I_{g,N_a}(t, i) &= \alpha_{N_a-1} I_{g,N_a-1}(t, i) - (\mu_{N_a}(t) + \nu) I_{g,N_a}(t, i) \\ &\quad + \beta(i) \sigma_{g,N_a} S_{g,N_a}(t) \phi_{N_a}(t, I)\end{aligned}\tag{6}$$

with $J_{g,a}^X(t)$: inflow of bought in ewes ($X = S, I$),
 $b(t, P)$: birth rate,
 $(G_{gg'})(t)$: breeding matrix,
 α_a : ageing coefficient,
 $\mu_a(t)$: reform rate,
 ν : mortality rate,
 $\sigma_{g,a}$: relative susceptibility ($\sigma_{g,a} = \sigma_g \sigma_a$),
 c_g : exponential growth rate of the infection load ($\frac{di}{dt} = c_g i$),
 $\beta(i)$: first infection load distribution (Beta distribution),
 $\phi_a(t, I)$: exposure \times force of infection.

$$\phi_a(t, I) = k_r I(t) + k \sum_{z=b,l} s_z(t) e_z(a) I_z(t)$$

with k_r : scaling factor for residual transmission (all year round),
 k : scaling factor for seasonal transmission (during lambing periods, $k \gg k_r$),
 z : $z = b$ for the main and $z = l$ the ewe-lamb lambing season,
 $s_z(t)$: seasonal function ($s_z(t) = 1$ if $t \in$ season z),
 $e_z(a)$: exposure coefficient during lambing season z ,
 $I_z(t)$: number of infected ewes that are lambing during season z .

Cinquième partie

Conclusion

Conclusion

Ce projet de thèse consistait à proposer des approches paramétriques permettant de modéliser et d’optimiser une sélection pour la résistance à une maladie en créant une population essentiellement porteuse des allèles de résistance au(x) gène(s) sélectionné(s) tout en minimisant l’impact de cette sélection sur le gain génétique acquis pour les caractères traditionnels de production. Un modèle déterministe et un modèle stochastique ont été développés pour atteindre cet objectif. Ces modèles permettent d’obtenir des stratégies optimales et d’évaluer différents scénarios (organisation du génotypage pour le gène de résistance dans la population sélectionnée, choix des conjoints selon leur génotype, méthode de diffusion des reproducteurs vers les élevages de production).

Le modèle déterministe qui a été développé permet d’optimiser simultanément la sélection sur deux caractères indépendants, un caractère monogénique pour la résistance aux maladies et un caractère de production polygénique. Il permet de s’adapter à de nombreuses situations puisqu’il prend en compte différents types de sélection (phénotypique, génotypique ou combinée génotype-phénotype) quelque soit le sexe (mâle et femelle) et différents types d’accouplements (raisonnés ou aléatoires).

Cette modélisation permet de concevoir des stratégies complexes mais également de voir si certaines simplifications sont envisageables afin de pouvoir les appliquer aisément en exploitation.

Ce modèle a été testé avec l’exemple de la tremblante. Cette méthodologie peut être extrapolée à la modélisation de la sélection de gènes affectant des caractères d’intérêt zootannique. Cette approche peut être étendue à plusieurs gènes/QTLs en intégrant également l’effet d’épistasie. Néanmoins ces évolutions seront limitées à cause d’une optimisation de plus en plus difficile due au nombre des paramètres augmentant de manière exponentielle.

Le modèle stochastique, qui a été développé lors de ce travail, permet de prendre en compte des fluctuations de la réponse à la sélection dues à des effectifs limités. Cette prise en compte est particulièrement importante à considérer quand la sélection se fait sur un gène identifié. L'optimisation d'un plan de sélection assisté par gène, quand le modèle prend en compte ces fluctuations, a soulevé néanmoins un certain nombre de problèmes spécifiques (méconnaissance des répercussions exactes des prises de décisions sur l'évolution du système en raison de la non unicité de la réponse à une décision prise ultérieurement, le nombre de variables important entraînant une modélisation complexe et une optimisation coûteuse en temps de calcul). C'est pourquoi nous nous sommes limités à un exemple simple afin de montrer l'utilité et la faisabilité de considérer un modèle stochastique. Néanmoins, il serait intéressant de simuler des situations plus complexes.

Les modèles développés lors de cette thèse peuvent être extrapolés à la modélisation de la sélection de gènes affectant des caractères d'intérêt zootechnique tel que la prolificité des brebis ou la conformation des carcasses d'agneaux. Concernant la prolificité, deux gènes majeurs ont déjà été mis en évidence dans la population Lacaune grâce au soutien financier de la région Midi-Pyrénées (Commission agriculture en 1996 ; Commission Recherche en 2000). Il serait donc intéressant d'optimiser la gestion de ces gènes au niveau de la population.

Le modèle déterministe développé a été couplé avec un modèle dynamique épidémiologique. La combinaison de ces deux modèles a permis d'étudier l'impact de différentes stratégies de sélection sur l'évolution de la tremblante. Le modèle épidémiologique a été développé au sein du laboratoire de Mathématiques et Informatique Appliquées de l'Institut national de la recherche Agronomique (INRA) par Suzanne Touzeau et a pour objectif d'étudier la propagation de la tremblante classique au sein d'un troupeau ovin. Ce travail a pu être réalisé à l'aide de données réelles issues de 3 races françaises : Lacaune, Manech Tête Rousse et Manech Tête Noire. Cette approche a permis de montrer que les stratégies de sélection avaient un fort impact sur l'épidémie de la maladie et étaient donc un outil efficace pour le contrôle de la tremblante. Ce travail a permis de mettre en relation des modèles développés en génétique avec des modèles développés en épidémiologie.

La connaissance du génome des différentes espèces est de plus en plus approfondies ouvrant de perspectives nouvelles pour la sélection pour la résistance aux maladies. Bishop et al (2007) ont mis en évidence que la variabilité de la sensibilité à de nombreuses maladies infectieuses avait une composante génétique. Ce constat permet d'envisager une sélection pour accroître la résistance. Néanmoins, la sélection doit prendre en compte les attentes des éleveurs. La sélection pour l'amélioration de la résistance aux maladies doit prendre en compte la sélection pour les autres caractères importants économiquement (production, fertilité,...). La valeur économique due à l'augmentation de la production est prioritaire comparé aux coûts occasionnés pour la lutte contre les maladies et les pertes qu'elles occasionnent (mort, coût des traitements). Une solution serait de créer un indice de sélection prenant en compte les relations génétiques entre les caractères pondérés par un poids économique qui permettrait de contrebalancer l'effet négatif de cette sélection sur les caractères de production.

Bibliographie

- Ackley, D. (1987), ‘An empirical study of bit vector function optimization. in : Genetic algorithms and simulated annealing, 170-204’.
- AFSSA (n.d.), ‘Afssa : Avis de l’afssa sur l’évolution des mesures de police sanitaire dans les cheptels ovins et caprins où un cas de tremblante classique ou atypique a été détecté. 2007’, pp. 1–9.
- Agerholm, J., Bendixen, C., Andersen, O. & Arnbjerg, J. (2001), ‘Complex vertebral malformation in holstein calves.’, *J. Vet. Diagn. Invest.* **13**, 283–289.
- Bacon, L. (1987), ‘Influence of the major histocompatibility complex on disease resistance and productivity’, *Poultry Science* **66**(5), 802–811.
- Baker, J. (1985), Adaptive selection methods for genetic algorithms, in ‘Proceedings of International Conference on Genetic Algorithms and Their Applications, Pittsburg, USA, 24-29 juillet 1985, 100-111’.
- Baker, J. (1987), Reducing bias and inefficiency in the selection algorithm, in ‘Proceedings of the 2nd International Conference on Genetic Algorithms, Massachusetts, USA, 28-31 Juillet 1987, 14-21’.
- Barillet, F., Rupp, R., Mignon-Grasteau, S., Astruc, J. & Jacquin, M. (2001), ‘Genetic analysis for mastitis resistance and milk somatic cell score in french lacaune dairy sheep.’, *Genet. Sel. Evol* **33**, 397–415.
- Belt, P., Muileman, I., Schreuder, B., Bos-de Ruijter, J., Gielkens, A. & Smits, M. (1995), ‘Identification of five allelic variants of the sheep prp gene and their association with natural scrapie’, *Journal of Genetic and Virology* **76**(3), 509–517.
- Benestad, S., Sarradin, P., Thu, B., Schonheit, J., Tranulis, M. & Bratberg, B. (2003), ‘Cases of scrapie with unusual features in norway and designation of a new type, nor98’, *Veterinary Record* **153**, 202–208.
- Bishop, S. & Woolliams, J. (2004), ‘Genetic approaches and technologies for improving the sustainability of livestock production’, *Journal of the Science of Food and Agriculture* **84**(9), 911–919.
- Booker, L. (1982), Intelligent bahavior as an adaptation to the task environment. doctoral dissertation, technical report n°243, ann arbor : University of michigan, logic of computers group. dissertation abstracts international, 43(2), 469b. 1-184, Technical report.

- Bouamama, S. & Ghédira, K. (2005), D3 g2a : A dynamic distributed double guided genetic algorithm for optimization and constraint satisfaction, *in* '17th International Conference on Parallel and Distributed Computing and System (IASTED-PDCS), Pheonix, Arizona, USA, 14-16 novembre 2005, 379-384'.
- Brindle, A. (1981), Genetic algorithms for function optimization. doctoral dissertation and technical report tr81-2. edmonton : University of alberta, department of computer science, Technical report.
- Brochard, M. & Palhiere, I. (2004), Side effects on genetic merit of the french breeding programme for scrapie resistance in berrichon du cher, causes du lot and mouton charollais breeds, *in* 'Rencontre Recherche Ruminants', Paris.
- Brochard, M., Palhière, I., Astruc, J., Barillet, F., Bouix, J., Bibe, B., Dion, F., Elsen, J., François, D., Griffon, L., Tiphine, L. & Tribon, P. (2006), Use of the prp major gene to select for scrapie resistance : example of 5 years intensive french breeding plan, *in* 'Proceedings of the 8th World Congress on Genetics Applied to Livest. Prod.', Vol. CD-Rom, comm. No. 22-32., Bello Horizonte, Brazil.
- Bubak, M., Ciesla, W. & Sowa, K. (1997), *Object-oriented library of parallel genetic algorithms and its implementation on workstations and HP/Convex Exemplar. In : High-Performance Computing and Networking, Springer, 514-523.*
- Buschmann, A., Biacabe, A., Ziegler, U., Bencsik, A., Madec, J., Erhardt, G., Lühken, G., Baron, T. & Groschup, M. (2004), 'Atypical scrapie cases in germany and france are identified by discrepant reaction patterns in bse rapid tests', *Journal of Virological Methods* **117**, 27-36.
- Carlborg, O., Andersson, L. & Kinghorn, B. (2000), 'The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci', *Genetics* **155**(4), 2003-2010.
- CERF, R. (1994), Une théorie asymptotique des algorithmes génétiques, PhD thesis, Montpellier II.
- Charlier, C., Coppieters, W., Rol, n. F. & Georges, M. (2008), 'Highly effective snp-based association mapping and management of recessive defects in livestock.', *Nat. Genet.* **40**, 449 :454.
- Cheng, H. (2005), Integrated genomic approaches to understanding resistance to marek s disease, *in* e. S.J. Lamont, M.F. Rothschild et D.L. Harris, ed., 'Proceedings of the third International Symposium on Genetics of Animal Health'.
- Cloucard, C., Beaudry, P., Elsen, J., Milan, D., Dussaucy, M., Bounneau, C., Schelcher, F., Chatelain, J., Launay, J. & Laplanche, J. (1995), 'Different allelic effects of the codons 136 and 171 of the prion protein gene in sheep with natural scrapie', *Journal of Genetic and Virology* **76**, 2097-2101.
- Colleau, J. (1999), Début de la sélection assistée par marqueurs chez les bovins laitiers : étude par simulation, *in* '6èmes Rencontres Recherches Ruminants, Paris, France, 1er et 2 décembre 1999, 231-233'.

- Colleau, J., Moureaux, S. & Briend, J. (2004), 'A method for the dynamic management of variability in dairy cattle', *Genetics Selection Evolution* **36**, 373–394.
- Colleau, J. & Tribout, T. (2008), 'Optimized management of genetic variability in selected pig populations', *Journal of Animal Breeding and Genetics* **125**(5), 291–300.
- Collet, C., Michou, L. & Audran, M. (2007), 'Paget's disease of bone in the french population novel sqstm1 mutations, functional analysis, and genotype phenotype correlations', *J.Bone Miner. Res.* **22**, 310–317.
- Comber, T. (1772), 'Real improvements in agriculture (on the principles of a. young, esq). letters to reade peacock, esq. and to dr hunter, physician in york, concerning the rickets in sheep', *Nicoll, London* pp. 73–83.
- Cuille, J. & Chelle, P. (1936), 'Pathologie animale. la maladie dite de la tremblante du mouton est-elle inoculable?', *Comptes Rendus de l'Académie des Sciences* (26), 1552–1554.
- Darwin, C. (1859), *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*, London John Murray, 321p.
- Davies, D. & Kimberlin, R. (1985), 'Selection of swaledale sheep of reduced susceptibility to experimental scrapie', *Veterinary Record* **116**, 211–214.
- Davis, L. (1987), *Genetic algorithms and simulated annealing*, Morgan Kaufmann Pub, 216p.
- Davis, T. E. & Principe, J. C. (1991), A simulated annealing like convergence theory for the simple genetic algorithm, in R. K. B. Booker & L. B., eds, 'Proceedings of the Fourth International Conference on Genetic Algorithms', Morgan Kauffman, San Mateo, CA., pp. 174–181.
- De Jong, K. (1980), 'Adaptive systeme design : A genetic approach', *IEEE Transactions on Systems, Man, and Cybernetics* **10**(3), 556–574.
- Dekkers, J. & Van Arendonk, J. (1998), 'Optimizing selection for quantitative traits with information on an identified locus in outbred populations', *Genetical Research* **71**, 257–275.
- Dickinson, A., Meikle, V. & Fraser, H. (1968), 'Identification of a gene which controls the incubation period of some strains of scrapie agent in mice', *Journal of Comparative Pathology* **78**, 293–299.
- Dorigo, M. & Gambardella, L. (1997), 'A cooperative learning approach to the traveling salesman problem', *IEEE Transactions on Evolutionary Computation, Man, and Cybernetics* **1**(1), 53–66.
- Edfors, L. & Wallgren, P. (2000), Escherichia coli and salmonella diarrhoea in pigs, in J. O. e. F. N. R.F.E. Axford, S.C. Bishop, ed., 'Breeding for resistance in Farm Animals', Wallingford, Royaume Uni., pp. 253–267.
- Elsen, J., Amigues, Y., Schelcher, F., Ducrocq, V., Andréoletti, O., Eychenne, F., Vu Tien Khang, J., Poivey, J., Lantier, F. & Laplanche, J. (1999), 'Genetic susceptibility and transmission factors in scrapie : detailed analysis of an epidemic in a closed flock of romanov', *Archives of Virology* **144**, 431–445.

- Everett, H. (1963), 'Generalized lagrange multiplier method for solving problems of optimum allocation of ressources', *Operations Research* **11**(3), 399–417.
- Fanen, P. & Hasnain, A. (2001), 'Mucoviscidose et gène cftr', *Atlas Genet Cytogenet Oncol Haematol*.
- Forrest, S. & Mitchell, M. (1993), 'Relative building-block fitness and the building-block hypothesis', *Foundations of Genetic Programming 2*.
- Foster, J., Hope, J. & Fraser, H. (1993), 'Transmission of bovine spongiform encephalopathy to sheep and goats', *Vet. Rec.*, **133**, 339–341.
- Fritz, S., Colleau, J., Druet, T., Boscher, M., Rossignol, M., Malafosse, A. & Boichard, D. (2003), Mise en place d'une sélection assistée par marqueurs dans les trois principales races bovines laitières françaises, in '10èmes Rencontres Recherches Ruminants, Paris, France, 3-4 décembre 2003, 53-57'.
- Gavier-Wide'n, D., Nöremark, M., Benestad, S., Simmons, M., Renström, L., Bratberg, B., Elvander, M. & Af Segerstad, C. (2004), 'Recognition of the nor98 variant of scrapie in the swedish sheep population', *Journal of Veterinary Diagnostic Investigation* **16**, 562–567.
- Gensure RC, e. a. (2005), 'A novel colla1 mutation in infantile cortical hyperostosis (caffey disease) expands the spectrum of collagen-related disorders', *J. Clin. Invest.* **115**, 1250 :1257.
- Gerardi, A. (1996), 'Bovine leukocyte adhesion deficiency : a review of a modern disease and its implications', *Research Veterinary Science* **61**, 183–186.
- Ghédira, K. & Jlifi, B. (2002), 'A distributed guided genetic algorithm for max-csps', *Revue d'Intelligence Artificielle* **16**(3), 367–382.
- Ghetti, B., Hutton, M. & Wszolek, Z. (2003), 'Frontotemporal dementia and parkinsonism linked to chromosome 17 associated with tau gene mutations (ftdp-17)', *Neurodegeneration : The Molecular Pathology of Dementia and Movement Disorders* pp. 86–102.
- Gibbons, R., Sellwood, R., Burrows, M. & Hunter, P. (1977), 'Inheritance of resistance to neonatal ecoli diarrhoea in the pig : examination of the genetic system.', *Theor. Appl Genet* **51**, 65–70.
- Gibson, J. (1994), short term gain at the expense of long term response with selection of identified loci, in 'Proceedings of the 5th world congress of genetics applied to livestock production', Vol. 21, pp. 201–204.
- Godi, G., Sauleau, R. & Thouroude, D. (2005), Conception d'Antennes Lentilles par Algorithme Génétique, in 'MajecSTIC 2005 : Manifestation des Jeunes Chercheurs francophones dans les domaines des STIC', pp. 380–384.
URL: <http://hal.inria.fr/inria-00000734/en/>
- Goldberg, D. (1989), *Genetics algorithms in search, optimization and machine learning*, Addison-Wesley Publishing Company, Inc., Boston.
- Goldberg, D. & Richardson, J. (1987), *Genetic algorithms with sharing for multimodal function optimization. In : Genetic Algorithms and their Applications*, Hillsdale, New Jersey and Lawrence Erl-baum Associates, 41-49.

- Goldmann, W., Hunter, N., Benson, G., Foster, J. & Hope, J. (1991), 'Different scrapie-associated fibril proteins (prp) are encoded by lines of sheep selected for different alleles of the sip gene', *Journal of Genetic and Virology* **72**, 2411–2417.
- Gong, Y., Slee, R. & Fukai, N. (2001), 'Ldl-receptor related protein 5 (lrp5) affects bone accrual and eye development', *Cell* **107**, 513–523.
- Grefenstette, J. (1981), Parallel adaptive algorithms for function optimization, Technical report.
- Grefenstette, J. & Baker, J. (1989), How genetic algorithms work : A critical look at implicit parallelism, in 'Proceedings of the 3rd International Conference on Genetic Algorithms, Virginie, USA, juin 1989, 20-27'.
- Grundy, B., Villanueva, B. & Woolliams, J. (1998), 'Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development', *Genetical Research* **72**, 159–168.
- Guaguère, E., Capt, A., Spirito, F. & Meneguzzi, G. (2003), 'L'epidermolyse bulleuse jonctionnelle du braque allemand : un modèle canin spontané de l'epidermolyse bulleuse jonctionnelle de l'homme', *Bulletin de l'Académie Vétérinaire de France* **156**(2), 51–55.
- Haldane, J. (1932), *The causes of evolution. Princeton Science Library Edition, Longmans, Green and CO, London England, 235pp.*
- Hansen, M., Van Zandt, J. & Law, G. (1967), 'Differences in susceptibility to marek's disease in chickens carrying two different b locus blood group alleles', *Poult. Sci.* .
- Hinrichs, D., Wetten, M. & Meuwissen, T. (2006), 'An algorithm to compute optimal genetic contributions in selection programs with large numbers of candidates', *Journal of Animal Science* **84**(12), 3212–3218.
- Holland, J. (1962a), *Concerning efficient adaptive systems. In M.C. Yovits, G.T. Jacobi, G.D. Goldstein (Eds), Self organizing systems (pp.215-230). Washinton : Spartan Books.*
- Holland, J. (1962b), 'Outline for a logical theory of adaptive systems', *J. ACM* **9**(3), 297–314.
- Holland, J. (1975), *Adaptation in Natural and Artificial Systems*, The MIT Press, 228 p.
- Hopfield, J. (1982), 'Neural networks and physical systems with emergent collective computational abilities', *Proceedings of the National Academy of Science USA* **79**(8), 2554–2558.
- Hospital, F. & Charcosset, A. (1997), 'Marker-assisted introgression of quantitative trait loci', *Genetics* **147**, 1469–1485.
- Hospital, F., Chevalet, C. & Mulsant, P. (1992), 'Using markers in gene introgression breeding programs', *Genetics* **132**, 1199–1210.
- Hunter, N., Foster, J., Goldmann, W., Stear, M., Hope, J. & Bostock, C. (1996), 'Natural scrapie in a closed flock of cheviot sheep occurs only in specific prp genotypes', *Archives of Virology* **141**, 809–824.

- Jeffrey, M., Martin, S., Gonzalez, L., Ryder, S., Bellworthy, S. & Jackman, R. (2001), 'Differential diagnosis of infections with the bovine spongiform encephalopathy (esb) and scrapie agents in sheep', *J. of Comp. Pathol.* **125**, 271–284.
- Joutel, A., Corpechot, C. & Ducros, A. (1996), 'Notch3 mutations in cadasil, a hereditary adult-onset condition causing stroke and dementia', *Nature* **383**(6602), 707–710.
- Karmarkar, N. (1984), 'A new polynomial time algorithm for linear programming', *Combinatorica* **4**(4), 373–395.
- Kijas, J., Bauer, T., Gafvert, S., Marklund, S., Trowald-Wigh, G., Johannisson, A., Hedhammar, A., Binns, M., Juneja, R., Hickstein, D. & Andersson, L. (1999), 'A missense mutation in the beta-2 integrin gene (itgb2) causes canine leukocyte adhesion deficiency', *Genomics* **61**, 101–107.
- Kijas, J., Juneja, R., Gafvert, S. & Andersson, L. (2000), 'Detection of the causal mutation for canine leukocyte adhesion deficiency (clad) using pyrosequencing', *Animal Genetic* **31**, 326–328.
- Kirkpatrick, S., Gelatt, C. & Vecchi, M. (1983), 'Optimization by simulated annealing', *Science* **220**(4598), 671–680.
- Kolehmainen, J., Black, G., Saarinen, A., Chandler, K., Clayton Smith, J., Traskelin, A., R., P., Kivitie-Kallio, S., Norio, R., Warburg, M., Fryns, J., de la Chapelle, A. & Lehesjoki, A. (2003), 'Cohen syndrome is caused by mutations in a novel gene, coh1, encoding a transmembrane protein with a presumed role in vesicle-mediated sorting and intracellular protein transport', *Am J Hum Genet* **72**, 1359–1369.
- Konold, T., Davis, A., Bone, G., Bracegirdle, J., Everitt, S., Chaplin, M., Saunders, G., Cawthraw, S. & Simmons, M. (2007), 'Clinical findings in two cases of atypical scrapie in sheep : a case report', *BMC Veterinary Research* **3**(2), 1–6.
- Konold, T., Davis, A., Bone, G., Simmons, M., Kahn, J., Blacke-Dyke, M., Bracegirdle, J. & Shimwell, C. (2006), 'Atypical scrapie cases in the uk', *Veterinary Record* **158**, 280.
- Krishnakumar, K. & Goldberg, D. (1992), 'Control system optimization using genetic algorithm', *Journal of Guidance, Control and Dynamics* **15**(3), 735–740.
- Lande, R. & Thompson, R. (1990), 'Efficiency marker assisted selection in the improvement of quantitative traits', *Genetics* **124**, 743–756.
- Laurens, J. (1995), Parallel Genetic Algorithms, PhD thesis.
- Le Dur, A. (2005), 'A newly identified type of scrapie agent can naturally infect sheep with resistant prp genotypes', *PNAS* **102**(44), 16031–16036.
- Lerman, I. & Ngouenet, F. (1995), Algorithmes génétiques séquentiels et parallèles pour une représentation affine des proximités. Rapport de Recherche de l'INRIA Rennes - Projet REPCO 2570, INRIA, PhD thesis.
- Lewis, F. (1986), *Optimal control*, Wiley, New York.

- Li, A., Jiao, X., Munier, F., Schorderet, D., Yao, W., Iwata, F., Hayakawa, M., Kanai, A., Shy Chen, M., Alan Lewis, R., Heckenlively, J., Weleber, R., Traboulsi, E. & Zhang, Q. (2004), 'Bietti crystalline corneoretinal dystrophy is caused by mutations in the novel gene *cyp4v2*', *Am J Hum Genet* **74**, 817–826.
- Ljungberg, K., Holmgren, S. & Carlborg, O. (2004), 'Simultaneous search for multiple quantitative trait loci using the global optimization algorithm direct', *Bioinformatics* **20**(12), 1887–1895.
- Luhken, G., Buschmann, A., Brandt, H., Eiden, M., Groschup, M. & Erhardt, G. (2007), 'Epidemiological and genetical differences between classical and atypical scrapie cases', *Veterinary Research* **38**, 65–80.
- Lush, J. (1937), *Animal Breeding Plan. The Iowa State College Press Edition*, 437pp.
- Mahfoud, S. & Goldberg, D. (1995), 'Parallel recombinative simulated annealing : A genetic algorithms', *Parallel Computing* **21**(1), 1–28.
- Maillard, J. C., Berthier, D., Chantal, I., Thevenon, S., Sidibe, I., Stachurski, F., Belem-saga, D., Razafindraibe, H. & Elsen, J. M. (2003), 'Selection assisted by a bola-dr/dq haplotype against susceptibility to bovine dermatophilosis', *Genetics Selection Evolution* **35**(Supl.1), S193–S200. Suppl. 1.
- Manfredi, E., Barbieri, M., Fournet, F. & Elsen, J. (1998), 'A dynamic deterministic model to evaluate breeding strategies under mixed inheritance', *Genetics Selection Evolution* **30**, 127–148.
- Marco, N., Godart, C., Désidéri, J., Mantel, B. & Périaux, J. (1996), A genetic algorithm compared with a gradient-based method for the solution of an active-control model problems. Technical report, INRIA, Rapport de Recherche de l'INRIA, Projet SINUS, n°2948, PhD thesis.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953), 'Equation of state calculation by fast computing machines', *Journal of Chemical Physics* **21**(6), 1087–1092.
- Meuwissen, T. (1997), 'Maximizing the response of selection with a predefined rate of inbreeding', *Journal of Animal Science* **75**, 934–940.
- Meuwissen, T. & Goddard, M. (1996), 'The use of marker haplotypes in animal breeding schemes', *Genet. Sel. Evol.* **28**, 161–176.
- Michalewicz, Z., Janikov, C. & Krawczyk, J. (1992), 'A modified genetic algorithm for optimal control problems', *Computers and Mathematics with Applications* **23**(12), 83–94.
- Milenkovic, D., Chaffaux, S., Taourit, S. & Guérin, G. (2003), 'A mutation in the *lamc2* gene causes the herlitz junctional epidermolysis bullosa (h-jeb) in two french draft horse breeds', *Genetics Selection Evolution* **35**, 249–256.
- Minvielle, F. (1998), *La Selection Animale. Collection Encyclopédique "Que sais-je"*, Presse Universitaire de France, Paris, 128pp.

- Montana, D. & Davis, L. (1989), Training feedforward networks using genetic algorithms, in 'Proceedings of the 11th International Joint Conference on Artificial Intelligence', Detroit, pp. 762–767.
- Morris, C., Vlassoff, A., Bisset, S., Baker, R., Watson, T., West, C. & Wheeler, M. (2000), 'Continued selection of romney sheep for resistance or susceptibility to nematode infection : estimates of direct and correlated responses', *Animal Science* **70**(1), 17–27.
- Moum, T., Olsaker, I., Hopp, P., Moldal, T., Valheim, M., Moum, T. & Benestad, S. (2005), 'Polymorphisms at codons 141 and 154 in the ovine prion protein gene are associated with scrapie nor98 cases', *Journal of Genetic and Virology* **86**(1), 231–235.
- Muhlenbein, H. & Schlierkaamp-Voosen, D. (1993), 'Predictive models for the breeder genetic algorithm : I. continuous parameter optimization', *Evolutionary Computation* **1**(1), 25–49.
- Nakabayashi, K., Amann, D. & Ren, Y. (2005), 'Identification of c7orf11 (tttn1) gene mutations and genetic heterogeneity in nonphotosensitive trichothiodystrophy', *Am. J. Hum. Genet.* **76**, 510 :516.
- Nakamichi, R., Ukai, Y. & Kishino, H. (2001), 'Detection for closely linked multiple quantitative trait loci using a genetic algorithm', *Genetics* **158**(1), 463–475.
- Nesterov, Y. & Nemirovski, A. (1994), 'Interior-point polynomial algorithms in convex programming', *SIAM Studies in Applied Mathematics Philadelphia* .
- Niaudet, P. & Gubler, M. (2007), 'Wt1 and glomerular diseases', *Pediatr. Nephrol.* **21**, 1653 :1660.
- Nix, A. & Vose, M. (1992), 'Modeling genetic algorithms with markov chains.', *Ann. Math. Artificial Intelligence* **5**, 79–88.
- Notarnicola, C., Didelot, M. N., Koné-Paut, I., Seguret, F., Demaille, J. & Touitou, I. (2002), 'Reduced mefv messenger rna expression in patients with familial mediterranean fever', *Arthritis and Rheumatism* **46**(10), 2785 :2793.
- Ollivier, L. (1998), 'The accuracy of marker-assisted selection for quantitative traits within populations in linkage equilibrium.', *Genetics* **148**, 1367–1372.
- Onnasch, H., Gunn, H., Bradshaw, B., Benestad, S. & Bassett, H. (2004), 'Two irish cases of scrapie resembling nor98', *Veterinary Record* **155**, 636–637.
- Orge, L., Galo, A., Machado, C., Lima, C., Ochoa, C., Silva, J., Ramos, M. & Simas, J. (2004), 'Identification of putative atypical scrapie in sheep in portugal', *Journal of Genetic and Virology* **85**(11), 3487–3491.
- Oulladji, L., Janka, A., Désidéri, J. & Dervieux, A. (2003), Optimisation aérodynamique par algorithmes génétiques hybrides : application à la réduction d'un critère de bang sonique, Technical report.
- Palhière, I., François, D., Elsen, J. M., Barillet, F. & Bouix, J. (2002), 'Alleles frequencies of the prp gene in 29 french sheep breeds. possible use in selection for resistance to scrapie', *Proceedings of the seventh world congress on genetics applied to livestock production* .

- Pereira, R. (2000), *Genetic algorithm optimisation for finance and investment*. In : *Munich Personal RePEc Archive*, La trobe Universtiy, Serie A 00.02, 1-29.
- Pong-Wong, R. & Woolliams, J. (2007), 'Optimisation of contribution of candidate parents to maximise genetic gain and restricting inbreeding using semidefinite programming', *Genetics Selection Evolution* **39**(1), 3–25.
- Prusiner, S. (1982), 'Novel proteinaceous infectious particles cause scrapie', *Science* **216**, 136–144.
- Reeves, C., ed. (1995), *Modern Heuristic Techniques for Combinatorial Problems*, Mc Graw-Hill.
- Rossi, L. (2001), 'Spink1/psti mutations are associated with tropical pancreatitis in bangladesh : a preliminary report', *Pancreatology* **1**, 242–245.
- Ruane, J. & Colleau, J. (1995), 'Marker-assisted selection for genetic improvement of animal populations when a single qtl is marked', *Genetical Research* **66**, 78–83.
- Ruane, J. & Colleau, J. (1996), 'Marker-assisted selection for a sexlimited character in a nucleus breeding population', *Journal of Dairy Science* **79**, 1666–1678.
- Rupp, R. & Boichard, D. (2003), 'Genetics of resistance to mastitis in dairy cattle', *Veterinary Research* **34**(5), 671–688.
- Sarradin, P., Lantier, F. & Elsen, J. (1996), 'La tremblante des bovins', *Note Interne INRA-ESR and ENA* .
- Saït, S. & Youssef, H., eds (1999), *Iterative computer algorithms with applications in engineering*, IEEE Computer Society Press.
- Shook, G. . & Schutz, M. (1994), 'Selection on somatic cell score to improve resistance to mastitis in the united states.', *J. Dairy Sci.*, **77**, 648 :658.
- Shovlin, C., Guttmacher, A., Buscarini, E., Faughnan, M., Hyland, R., Westermann, C., Kjeldsen, A. & Plauchu, H. (2000), 'Diagnostic criteria for hereditary hemorrhagic telangiectasia', *Am. J. Med. Genet.* **91**, 66 :67.
- Shuster, D., Kehrli, M., Ackermann, M. & Gilbert, R. (1992), 'Identification and prevalence of a genetic defect that causes leukocytes adhesion deficiency in holstein cattle', *Proceeding National Academy of Science* **89**, 9225–9229.
- Siarry, P. & Dreyfus, G., eds (1989), *La méthode du recuit simulé : théorie et applications*, ESPCI-IDSET.
- Sigurdsson, B. (1954), 'Observations on three slow infections of sheep', *British Veterinary Journal* **110**, 225–270.
- Simmons, M., Konold, T., Simmons, H., Spencer, Y., Lockey, R., Spiropoulos, J., Everitt, S. & Clifford, D. (2007), 'Experimental transmission of atypical scrapie to sheep', *BMC Veterinart Research* **3**(20), 1–7.

- Smith, C. (1967), 'Improvement of metric traits through specific genetic loci.', *Animal Production* **9**, 349–358.
- Soller, M. & Medjugorac, I. (1999), A successful marriage : making the transition from quantitative trait locus mapping to marker assisted selection, *in* M. F. R. J. C. M. Dekkers, S. J. Lamont, ed., 'Jay L. Lush to genomics : visions for animal breeding and genetics'.
- Sonesson, A. & Meuwissen, T. (2000), 'Mating schemes for optimum contribution selection with constrained rates of inbreeding', *Genetics Selection Evolution* **32**(3), 231–248.
- Sonesson, A. & Meuwissen, T. (2002), 'Non-random mating for selection with restricted rates of inbreeding and overlapping generations', *Genetics Selection Evolution* **34**(1), 23–39.
- Spelman, R. & Van Arendonk, J. (1997), 'Effect of inaccurate parameter estimates on genetic response to marker assisted selection in an outbred population', *Journal of Dairy Science* **80**, 3399–3410.
- Talbi, E. & Bessière, P. (1991a), Genetic parallel algorithm : performances and applications, *in* 'Proceedings of International Conference on Novel Optimization Technics, Copenhagen, Denmark.'
- Talbi, E. & Bessière, P. (1991b), A parallel genetic algorithm for the graph partitioning problem, *in* 'A.C.M. International Conference on SuperComputing, Cologne, Germany'.
- Vallejo, R., Bacon, L., Liu, H., Witter, R., Groenen, M., Hillel, J. & Cheng, H. (1998), 'Genetic mapping of quantitative trait loci affecting susceptibility to Marek's disease induced tumours in F2 intercross chickens', *Genetics Selection Evolution* **148**(1), 349–360.
- Van der Beek, S., Van Arendonk, J. & Groen, A. (1995), 'Power of two- and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families', *Theoretical Applied Genetics* **91**, 1115–1124.
- Vandenberghe, L., Balakrishnan, V., Wallin, R. & Hansson, A. (2003), On the implementation of primal dual interior point methods for semidefinite programming problems derived from the KYP lemma, *in* 'Proceedings of the 42nd IEEE Conference on Decision and Control', p. 4658 :4663.
- Verrier, E. (2001), 'Marker-assisted selection for the improvement of antagonistic traits under mixed inheritance', *Genetics Selection Evolution* **33**, 17–38.
- Villanueva, B., Dekkers, J., Woolliams, J. & Settar, P. (2004), 'Maximizing genetic gain over multiple generations with quantitative trait locus selection and control of inbreeding', *Journal of Animal Science* **82**, 1305–1314.
- Villanueva, B., Pong-Wong, R. & Woolliams, J. (2002), 'Marker assisted selection with optimised contributions of the candidates to selection', *Genet. Sel. Evol.* **34**, 679–703.
- Visscher, P., Haley, C. & Thompson, R. (1996), 'Marker-assisted introgression in backcross breeding programs', *Genetics* **144**, 1923–1932.
- Vitezica, Z. G., Moreno, C. R., Bodin, L., François, D., Barillet, F. & Elsen, J. M. (2006), 'No associations between prp genotypes and reproduction traits in Inra 401 sheep.', *J Anim sci*

- Wells, G., Scott, A., Johnson, C., Gunning, R., Hancock, R., Jeffrey, M., Dawson, M. & Bradley, R. (1987), 'A novel progressive spongiform encephalopathy in cattle', *Veterinary Record* **121**, 419–420.
- Winter, G., Galan, M., Cuesta, P. & Greiner, D. (1995), *GAs in Engineering and Computer Science : GAs. In : A stochastic Improvement Technique, Tools, Skills, Pitfalls and Examples*, J.Wiley and Sons, N.Y., 217-250.
- Wisniewski, H., Sigurdarson, S., Rubenstein, R., Kassak, R. & Carp, R. (1996), 'Evidence from america and iceland that infection of scrapie can remain in infected hay mites', *The Lancet* **347**(9008), 1114.
- Woolaston, R. (1992), 'Selection of merino sheep for increased and decreased resistance to haemonchus contortus : peri-parturient effects on faecal egg counts', *International Journal for Parasitology* **22**(7), 947–953.
- Woolaston, R. & Windon, R. (2001), 'Selection of sheep for response to trichostrongylus colubriformis larvae : genetic parameters', *Animal Science* **73**(1), 41–48.
- Woolliams, J., Pong-Wong, R. & Villanueva, B. (2002), Strategic optimisation of shortand long-term gain and inbreeding in mas and non mas schemes, in '7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, 19-23 Aout 2002, Com.23-02, 155-162'.
- Wray, N. & Goddard, M. (1994), 'Increasing long term response to selection', *Genetics Selection Evolution* **26**, 431–459.
- Yonash, N., Bacon, L., Witter, R. & Cheng, H. (1999), 'High resolution mapping and identification of new quantitative trait loci (qtl) affecting susceptibility to marek s disease', *Animal Genetics* **30**(2), 126–135.
- Zhang, D., Herring, J. A., Swaney, S. S., McClendon, T. B., Gao, X., Browne, R. H., Rathjen, K. E., Johnston, C. E., Harris, S., Cain, N. M. & Wise, C. A. (2006), 'Mutations responsible for Larsen syndrome cluster in the FLNB protein', *Journal of Medical Genetics* **43**(5), e24.
URL: <http://jmg.bmj.com/content/43/5/e24.abstract>